2006-12

The many facets of Internet topology and traffic

Alderson, D.

American Institute of Mathematical Sciences

Neworks and Heterogenous Media, v. 1, no. 4, December 2006, pp. 569-600.

NETWORKS AND HETEROGENEOUS MEDIA
Website: http://aimSciences.org

Volume **1**, Number **4**, December **2006**
pp. **569–600**

THE MANY FACETS OF INTERNET TOPOLOGY AND TRAFFIC

D. ALDERSON
Operations Research Department,
Naval Postgraduate School, Monterey, CA 93943, USA

H. CHANG
Department of EECS, University of Michigan
Ann Arbor, MI 48109-2122, USA

M. ROUGHAN
School of Mathematical Sciences,
University of Adelaide, Adelaide 5005, Australia

S. UHLIG
Network Architectures and Services,
Delft University of Technology, Delft, The Netherlands

W. WILLINGER
AT&T Labs-Research, Florham Park, NJ 07932, USA

ABSTRACT. The Internet's layered architecture and organizational structure give rise to a number of different topologies, with the lower layers defining more physical and the higher layers more virtual/logical types of connectivity structures. These structures are very different, and successful Internet topology modeling requires annotating the nodes and edges of the corresponding graphs with information that reflects their network-intrinsic meaning. These structures also give rise to different representations of the traffic that traverses the heterogeneous Internet, and a traffic matrix is a compact and succinct description of the traffic exchanges between the nodes in a given connectivity structure. In this paper, we summarize recent advances in Internet research related to (i) inferring and modeling the router-level topologies of individual service providers (i.e., the physical connectivity structure of an ISP, where nodes are routers/switches and links represent physical connections), (ii) estimating the intra-AS traffic matrix when the AS's router-level topology and routing configuration are known, (iii) inferring and modeling the Internet's AS-level topology, and (iv) estimating the inter-AS traffic matrix. We will also discuss recent work on Internet connectivity structures that arise at the higher layers in the TCP/IP protocol stack and are more virtual and dynamic; e.g., overlay networks like the WWW graph, where nodes are web pages and edges represent existing hyperlinks, or P2P networks like Gnutella, where nodes represent peers and two peers are connected if they have an active network connection.

1. Introduction.

The design and implementation of most complex systems is inevitably broken down into simpler subsystems that tend to be separately optimized and implemented and then interconnected, often in an ad-hoc manner. A prime example of this approach is the architecture of the Internet, which is comprised of a modular design based on a *dual decomposition* of functionality–a vertical separation into layers and a horizontal decentralization across network components [121]. One of the most visible manifestations of the Internet's vertical decomposition is the 5-layer TCP/IP protocol stack, consisting of (from

2000 *Mathematics Subject Classification.* Primary: 90B15; Secondary: 90B10.
Key words and phrases. Internet topology, network tomography, traffic matrix, gravity model, model validation.

the bottom up) the physical layer (e.g., optical fiber, copper), the data link or network access layer (e.g., Ethernet, frame relay), network or internet layer (e.g., *Internet Protocol*, or IP), the host-to-host or transport layer (e.g., *Transmission Control Protocol*, or TCP), and the application layer (e.g., *HyperText Transfer Protocol*, or HTTP, for the World Wide Web, or WWW). Here, each layer hides the complexity of the layer below and provides a well-defined service to the layer above (e.g., congestion and error control in TCP, routing in IP).

Together, these layers support a reliable communication service over a large array of geographically dispersed hardware components (e.g., routers, switches, fiber) with different functionality, capability, reliability, and ownership. A manifestation of the Internet's horizontal decomposition is how this large-scale physical infrastructure is organized into *Autonomous Systems*. Here, an autonomous system (AS) or autonomous domain is a group of routers and networks managed by a single organization. In turn, an *Internet Service Provider (ISP)* can consist of a single AS or of a group of ASes, but for simplicity, we will use the terms AS and ISP indistinguishably throughout this paper. The TCP/IP protocol stack as a whole and IP in particular are able to hide from the user much of the enormous complexity associated with controlling this diverse set of networked resources and coordinating the actions among the many competing ISPs. By providing the mechanisms necessary to knit together diverse networking technologies and ASes into a single virtual network (i.e., a network of networks, or "Internet"), they ultimately guarantee seamless connectivity and reliable communication between sending and receiving hosts, irrespective of where in the network they are.

While this architecture of the Internet has enabled remarkable flexibility, extensibility, scalability, and robustness for the network as a whole, it has also created significant challenges for an exact characterization of network structure, behavior, and traffic. In particular, it makes the inspection of the Internet's topology and traffic difficult for two reasons. First, there does not exist a single vantage point from which one can "see" the entire network (i.e., there is no central authority). Second, because each layer of the architecture defines its own connectivity and is governed by its own protocol dynamics, the meaning of network "topology" and "traffic" depends directly on one's choice of focus. This problem is greatly compounded by a general reluctance on the part of the "owners" of the Internet (i.e., ISPs) to share detailed information about their networks due to proprietary reasons, to protect the privacy of their customers, and for fear of loosing their competitive advantage in the fiercely contested ISP market.

Due to a general lack of publicly available information about the infrastructure of the Internet and the traffic that it carries, networking researchers have been faced with the task of developing measurement and analysis tools of varying sophistication to reverse engineer the various structures and corresponding traffic flows. The general problem they have to deal with can be broken down into four distinct tasks: (1) *measurements*, (2) *inference*, (3) *modeling*, and (4) *model validation*. For each task, there are a number of challenges and difficulties, and considerable ambiguity exists with respect to all four of them, but typically for very different reasons. Yet, considerable progress has been made in these areas over the last decade, and it is the intent of this paper to provide the appropriate background and describe and illustrate the state-of-the-art for newcomers and nonspecialists with a number of concrete examples.

In the process of recounting some of the most notable contributions in these areas, we observe several recurring themes that appear in the literature and which we summarize here.

1. *Internet measurements are typically of varying quality.* They are often imperfect or incomplete and can contain errors or ambiguities that depend on the process by which they have been collected. In general, Internet measurements should not be taken at face value, but need to be scrutinized for consistency with the networking context from which they were collected.
2. *Inference from quantitative data is only as good as the data that underlies the inference process.* The challenge is to know whether or not *"the results we [infer] from our measurements are indeed well-justified claims"* [79], and at issue are the quality of the measurements themselves, the quality of their analysis, and the sensitivity of the inferred properties to measurement errors.
3. *Developing appropriate models that elucidate observed structure or behavior is typically an underconstrained problem*, meaning that there are in general many different explanations for one and the same phenomenon. To argue in favor of any particular explanation typically involves additional information, either in the form of domain knowledge or of new or complementary data. It is in the choice of this information and how it is incorporated into the model building process, where considerable differences arise in the various approaches that have been applied.
4. *There has been an increasing awareness of the fact that the ability to replicate some statistics of the original data or inferred quantities does not constitute validation for a particular model.* While one can always use a model with enough parameters to "fit" a given data set, such models are merely descriptive and have in general no explanatory power. For the problems described here, appropriate validation typically means identifying and collecting complementary measurements that enable the "closing the loop" in the research process in the sense of [120].

Before proceeding with our review of recent contributions, we note that a fundamental, yet open, question in the study of complex networks in general is, *What elements of a network are most important in a representative model?* Most researchers will readily agree that connectivity as defined by nodes and arcs is fundamental (i.e., they are the mathematical primitives of a graph), but most practical network models of the Internet require some type of *annotation* (e.g., bandwidth, delay) beyond simple connectivity. Moreover, for researchers involved in the analysis of complex networks across disciplines (e.g., biology, sociology—see [72] for a representative discussion), reducing the complex function of these diverse systems to a simple graph is often the only way to obtain a common denominator for comparison. As a result, considerably more effort has focused on the statistical properties of graphs in general than on their domain-specific structure and function. The behavior of networks, as defined by the dynamical processes that run on top of a given topology are even less understood, except in specific cases.

In what follows, we describe some of the main efforts and approaches to discovering and understanding Internet topology and traffic, as viewed from a number of different perspectives on the Internet. In particular, in Section 2 we take the perspective of a network researcher interested in reverse engineering the physical infrastructure of a single AS as well as its traffic. For the former, we assume no access to proprietary data, while for the latter, we require access to ISP-specific information. In Section 3 we consider the case of a researcher interested in reverse engineering the Internet's AS-level topology and the corresponding inter-AS traffic exchanges, without access to any proprietary, AS-specific data sources. Lastly, Section 4 gives a brief description of existing work on discovering the structure and evolution of overlay networks such as the Web graph and certain P2P networks. We conclude in Section 5 with a discussion of a number of challenging open problems and future research directions.

2. **The Case of a Single Autonomous System.** In this section, we consider a single AS and are interested in its router-level topology (i.e., the layout of the AS's physical infrastructure consisting of routers, switches, and physical cables) and in the traffic that traverses this infrastructure. In general, while the administrator of an AS has detailed knowledge about its physical infrastructure, he/she has no intrinsic means for knowing its traffic matrix. Since concern for customer privacy and fear of losing competitive advantage have provided a strong disincentive for network owners and operators to share topology information, direct inspection of an AS's network is generally not possible, and researchers have used both empirical and theoretic approaches to discover its physical infrastructure by probing the AS's network "from the outside." On the other hand, since obtaining accurate traffic matrices as a means for describing the traffic that traverses its own network is of considerable interest to an AS, we take the perspective of a researcher who has access to AS-specific information (e.g., router-level topology, routing matrix, link-load measurements) and is concerned with the problem of estimating the intra-domain traffic matrix for the AS in question.

2.1. **Intra-AS router-level topology.** For router-level related issues such as performance, reliability, and robustness to component loss, the physical connectivity between routers is more important than the virtual connectivity as defined by the higher layers of the protocol stack (e.g., IP). Thus, when referring in the following to router-level connectivity, we always mean the data link or network access layer (i.e., Layer 2), especially when the distinction between this layer and the network or internet layer (i.e. Layer 3) is important for the purpose of illuminating the nature of the actual router-level connectivity (e.g., node degree) and its physical constraints.

2.1.1. *Measurements.* Because most ISPs consider their router-level topology to be proprietary, the location and connectivity of routers within the Internet cannot be measured directly, and coaxing from them the quantities of interest typically requires significant effort and involves more or less sophisticated heuristics for interpreting any data obtained. For example, one-hop connectivity between routers running IP can be observed using `traceroute`, which records successive IP-hops along paths between selected network host computers [65]. Traceroute has been successfully incorporated into measurement experiments of varying scope (see for example, the Mercator [49], Skitter [32], Rocketfuel [96], and DIMES [92] projects), and it remains one of the most popular tools in use today for router-level mapping.

2.1.2. *Inferring intra-AS router-level connectivity.* Ongoing research continues to reveal more and more idiosyncrasies of traceroute data and shows that their interpretation requires great care and diligent mining of other available data sources. For example, a primary challenge in trying to reverse-engineer a network's physical infrastructure from traceroute-based measurements is that IP connectivity is an abstraction (at Layer 3) that sits on top of physical connectivity, so traceroute is unable to record directly the network's physical structure, and its measurements are highly ambiguous about the dependence between these two layers. Such ambiguity in Internet connectivity persists even at higher layers of the protocol stack, where connectivity becomes increasingly virtual, but for different reasons (for example, see below for a discussion of the Internet's AS and Web graphs).

The challenges associated with disambiguating the available (traceroute-based) measurements and identifying those contributions that are relevant for the Internet's router-level topology can be daunting. In particular, using traceroute measurements at face value and submitting them to commonly-used, black box-type modeling techniques has been problematic for three reasons.

1. Traceroute data often gives the false appearance of direct or high connectivity among routers. To illustrate how the somewhat subtle interactions among the different layers of the Internet protocol stack can give the (false) appearance of high connectivity at the IP-level, recall how at the physical layer the use of Ethernet technology near the network periphery or Asynchronous Transfer Mode (ATM) technology in the network core can give the appearance of high IP-connectivity since the physical topologies associated with these technologies may not be seen by IP-based traceroute. In such cases, machines that are connected to the same Ethernet or ATM network may have the illusion of direct connectivity from the perspective of IP, even though they are separated by an entire network (potentially spanning dozens of machines or hundreds of miles) at the physical level. See Figure 1 below for an example. In an entirely different fashion, the use of "Layer 2.5 technologies" such as Multiprotocol Label Switching (MPLS) tend to mask a network's physical infrastructure and can give the illusion of one-hop connectivity at Layer 3. Note that in both cases, it is the explicit and intended design of these technologies to hide the physical network connectivity from IP.

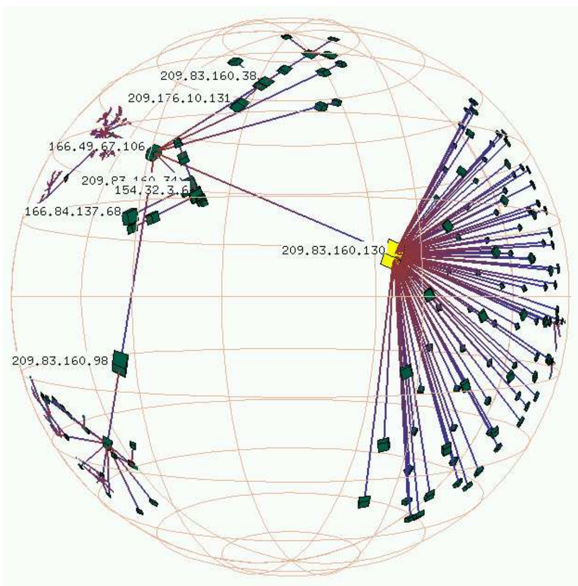


FIGURE 1. Sample visualization of IP-level topology from the Skitter Project [32] which uses traceroute to measure connectivity. The IP address `209.83.160.130` belongs to `savvis.net`, a managed IP and hosting company offering "private IP with ATM at core," so it is likely that in reality this highly connected "node" corresponds to an entire ATM network (at Level 2) and not just a router. (Image courtesy UCSD/CAIDA, ©The Regents of the University of California.)

2. It is nontrivial to determine which IP addresses belong to the same router. Because traceroute records only successive IP hops along a given path and there is typically a different IP address associated with each *physical interface card (PIC)* on a router, one must first decide which IP addresses/PICs (and corresponding DNS names) refer to the same router, a process known as *alias resolution* [95]. While one of the contributing factors to the high fidelity of the current state-of-the-art Rocketfuel maps is the use of an improved heuristic for performing alias resolution [96], further ambiguities and inaccuracies remain, as pointed out for example in [106, 8].

3. The experimental setup of some traceroute studies may introduce statistical biases that considerably alter the nature of the discovered maps. Recently, both experimental [57] and theoretical [1] research has identified a potential bias whereby IP-level connectivity in traceroute studies is inferred more easily and accurately the closer the routers are to the traceroute source(s). Such bias possibly results in incorrectly interpreting a network to have power law-type distributions in node *degree* (i.e., connectivity) when the true underlying connectivity structure is a regular graph (e.g., Erdös-Renyí [38]).

Because of these potential inaccuracies with traceroute data, the process by which one models and validates a candidate router-level topology is extremely important.

2.1.3. *Router-level topology modeling and model validation.* The development of abstract, yet informed, models for intra-AS network topology analysis and generation has leveraged both empirical and theoretical approaches. The first popular topology generator to be used for networking simulation was the Waxman model [119], which is a variation of the classical Erdös-Rényi random graph [38] in which nodes are placed at random in a two-dimensional plane and links are added probabilistically in a manner that is inversely proportional to their Euclidean distance. Although this model followed the general observation that long-range links in real networks are expensive, the use of this type of random graph model was later abandoned in favor of models that explicitly introduce non-random structure, particularly hierarchy and locality, as part of the network design [35, 18, 125]. The development of these *structural topology models* (in particular, the Georgia Tech Internetwork Topology Models or GT-ITM) was based on the fact that an inspection of real networks shows that they are clearly not random but do exhibit certain obvious hierarchical features. An important contribution of this work was the argument that a topology model (and generator) should reflect the design principles in common use. For example, in order to achieve desired performance objectives, the network must have certain connectivity and redundancy requirements, properties which are not guaranteed in random network topologies.

After the development of large-scale, traceroute-based measurement studies resulted in broad availability of IP-level connectivity data, the emphasis on Internet topology modeling and analysis shifted to the study of aggregate statistical properties and their explanations [40, 93, 49, 32, 96, 103]. Of primary interest within the literature have been network statistics related to the *connectivity* of network components, whether they be machines in the router-level graph or entire subnetworks in the AS-level graph (e.g., [48, 22]—see Section 3.1 for details). Consistent with a broader debate within the complex systems community, considerable attention has been devoted to the prevalence of heavy-tailed distributions in node *degree* (e.g., number of connections) and whether or not these heavy-tailed distributions conform to scaling (i.e., power-law) distributions [40, 93, 67, 27, 69].

Power laws in the network node connectivity have been a popular topic in the study of networks across disciplines, because this simple statistic captures in a parsimonious manner a prominent feature of many real-world networks, namely that most nodes have very few connections and a few nodes have lots of connections. This feature has been a central issue in the study of so-called *scale-free* network models [11], which have been a popular theme in the study of complex networks, particularly among researchers inspired by statistical physics [72, 6, 78]. However, in the study of Internet topology, the discovery of power laws from traceroute studies (ambiguities of traceroute measurements notwithstanding) have also greatly influenced the recent generation and evaluation of network models.

In the current environment, node degree distributions and other large-scale statistics are popular metrics for evaluating how representative a given topology is [104], and proposed topology generators are often evaluated on the basis of whether or not they can reproduce the same types of macroscopic statistics, especially power law-type node degree distributions [16]. Because the structural topology generators in GT-ITM fail to produce power laws in node degree, they were abandoned in favor of a new class of *degree-based* generators (see [61] for a partial list) that explicitly replicate these observed statistics. The popularity of these generators notwithstanding, this emphasis on power-law node degree distributions and the resulting efforts to generate and explain them with the help of newly developed models have been met with considerable criticism for several reasons.

1. The ambiguities inherent in traceroute-based studies suggest that the appearance of strict power laws should be viewed with healthy skepticism. As noted above, there are both statistical as well as measurement-based reasons why traceroute data may give the false appearance of high connectivity routers. Arguments in favor of degree-based generators often rely on their ability to match exactly the exponent of an observed power law [11]—an argument that breaks down if the real distribution differs from what has been inferred from measurements.

2. The degree distribution does not uniquely characterize a graph, the forces governing its structure, or its behavior. There is considerable diversity in the space of graphs having the same degree sequence, such that two graphs having the *same* power-law degree distribution may be viewed as *opposites* from an engineering perspective that incorporates router capacity and is motivated by network throughput [61, 62]. Furthermore, many graphs matching an observed power-law distribution may have no network-intrinsic meaning whatsoever or may be unrealizable from existing hardware [61, 8].

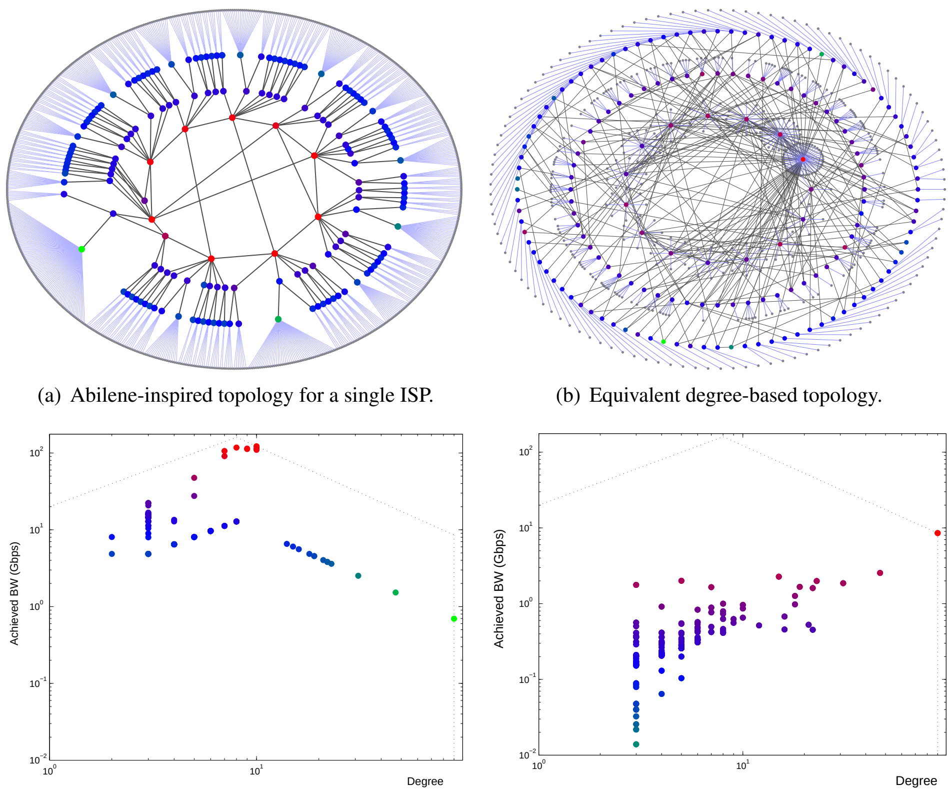
3. There remains considerable debate as to the significance, if any, for power law distributions in the node degree of a network. For example, there is a long-standing but little-known argument originally due to Mandelbrot [66, pp. 79–116] (see also [122, 123]) which says in short that power-law type distributions should be expected to arise ubiquitously for purely mathematical and statistical reasons. A detailed discussion of the debate surrounding power laws and "scale free" networks is available from [62, 37].

4. Degree-based models are merely descriptive and not explanatory. There are many ways of producing power laws [73], so a generative model that merely replicates an observed distribution provides no evidence of a correct physical explanation for the overall network structure [120]. In other words, without an understanding of the main drivers of network deployment and growth, it is difficult to identify the causal forces affecting large-scale network properties and even more difficult to predict future trends in network evolution.

A simple example is helpful in characterizing the potentially extreme differences between degree-based topologies and those inspired by engineering. Borrowing from the illustrative networks first presented in [61], Figure 2 shows two network topologies having the *same degree distribution*, which happens to be of the power-law type. The network in 2(a) was inspired by the Abilene educational backbone network (`http://abilene.internet2.edu/`), and the network in 2(b) was generated from a degree-based method (i.e., [33]). In comparing the functionality of these two networks, we define *network performance* as the maximum throughput on the network under heavy traffic conditions based on a gravity model (e.g., [54] with a detailed discussion in Section 2.2). That is, we consider flows on all source-destination pairs of edge routers, such that the amount of flow X_{ij} between source i and destination j is proportional to the product of the traffic demand x_i, x_j at end points i, j, $X_{ij} = \alpha x_i x_j$, where α is some constant. We compute the maximum throughput on the network as

$$\begin{aligned} \max_{\alpha} \qquad & \sum_{i,j} \alpha x_i x_j \\ s.t \qquad & R\mathbf{x} \le \mathbf{b}, \end{aligned} \tag{1}$$

where $\mathbf{x}$ is a vector obtained by stacking all the flows $X_{ij} = \alpha x_i x_j$ and R is the routing matrix (defined such that $R_{kl} = \{0, 1\}$ depending on whether or not flow l passes through router k). We use shortest path routing to get the routing matrix, and define $\mathbf{b}$ as the vector consisting of all router bandwidths according to the degree bandwidth constraint (i.e., the convex region in 2(c-d)—see [61] for a detailed discussion) which limits the allowable density of router degree-bandwidth. Under the same traffic assumptions and routing constraints, the network in 2(a) can carry more than 20 times the traffic of network 2(b). The cause of this discrepancy is easily seen in 2(c-d), which shows that under maxflow conditions the degree-based network in 2(b) suffers from severe bottlenecks, while the design in 2(a) does a better job in reconciling the routing and capacity tradeoffs.



(a) Abilene-inspired topology for a single ISP.

(b) Equivalent degree-based topology.

(c) Maxflow utilization of routers in (a).

(d) Maxflow utilization of routers in (b).

FIGURE 2. Differences between reverse-engineered and degree-based topologies. As first presented in [61], the networks in (a) and (b) have the same (power-law) degree distribution. Under maxflow conditions defined by Eq.(1), the network in (a) achieves 3.95×10^{11} Mbps total throughput while the network in (b) achieves only 1.64×10^{10} Mbps, a difference of more than $20\times$.

Although the deficiencies of degree-based methods are now well-understood, in the absence of concrete examples of alternate models these methods have remained popular representations for router-level Internet structure. In choosing random graphs (often constrained to match a particular statistic of interest, such as a degree distribution) as the starting point for a model, these approaches typically look for the "most likely" network configuration matching the chosen statistic. Recent attempts to overcome the problems associated with the degree distribution of a graph have focused on more sophisticated graph statistics, such as the *joint degree distribution (JDD)* or higher-order correlation structures [64], these approaches still suffer from the basic problem that the "most likely" configuration may not correspond to a semantically meaningful or functional network.

The development of router-level topology models that adhere to engineering reality and are consistent with the statistics of large-scale measurement studies remains an open area of research. However, recent work reported in [8] that builds on arguments in [61, 7] advocates an approach that is primarily based on reverse-engineering observed structure by reconciling the drivers of engineering design with the observed high variability in topology-related measurements. The basic tenets of this approach are the following.

- The principal decision makers in the context of network provisioning, design, and management are the ISPs, and the model should reflect the real objectives and constraints that they face. In essence, this approach returns to the basic perspective advocated by proponents of structural models [35, 18, 125].
- The construction and evolution of real networks is governed by a tradeoff between what is desirable (i.e., the objectives of the ISP) and what is possible (i.e., the constraints facing the ISP). This tension is naturally captured in the form of a mathematical program (i.e., a constrained optimization problem).
- The emphasis in modeling is on identifying a simple characterization of the most important objectives (e.g., throughput maximization) and constraints (e.g., router linecard capacity and density) that drive the decisions of the ISP. By choosing objectives and constraints that are consistent with the engineering details, one guarantees that any resulting model will adhere to reality. The ability to replicate aggregate statistics of observed networks is taken as secondary evidence only.
- Because the approach is easily extended to include new objectives or constraints, the approach naturally lends itself to a feedback loop by which one seeks out additional validation (e.g., in the form of new measurement studies) of the key objectives and constraints and then incorporates them into the modeling framework.

While there is considerable work to be done in the development of appropriate topology generators based on this optimization-based approach, simple models using this framework have provided proof-of-concept in the ability of simple models reflecting the appropriate engineering tradeoffs to provide superior explanations to measured statistics, including power laws [61, 8, 62]. Of course, new advances in the ability to measure directly the router-level structure of the Internet would go a long way to reducing the ambiguity and guesswork involved in topology modeling and generation.

2.2. **Intra-AS traffic matrices.** Topologies do not appear in a vacuum. They are, as noted, designed to carry traffic. Hence our understanding of a topology cannot be complete without some understanding of the traffic carried by that topology; that is, its *traffic matrix*. Here, a traffic matrix describes the volume of traffic flowing between pairs of points in a network. Traffic matrices are studied because they are expected to define an invariant in the sense that they are largely insensitive to changes of the topology. Hence, they can be used, for example, to compare different network topologies, and make judgments about which is optimal under some criteria.

Work on Internet traffic matrices started as far back as 1974 [53] with an analysis of the traffic distribution on the ARPANET, but was quite restricted until a decade ago when Vardi [113] coined the term "Network Tomography" for the problem of inferring traffic matrices from link load measurements. Vardi's work presented both an interesting statistical problem, and an easier method for obtaining traffic matrices, which are not intrinsically available to a network administrator. They must be measured (involving considerable effort, and network overhead), or inferred by tomography.

We can study traffic matrices at various granularities: e.g., computer to computer, router to router, or PoP (for Point-of-Presence) to PoP, and from varying points of focus. Vardi [113], Kleinrock [53] and much of the subsequent literature on traffic matrices were limited to observations of a single network, and therefore describe intra-AS or intra-domain traffic matrices. This section is concerned with these intra-domain traffic matrices, but it is important to be precise about the definition of intra-domain traffic matrix. Typical papers on traffic matrices refer to an Origin-Destination (OD) traffic matrix (though these are sometimes called Source/Destination traffic matrices or demands). Origin (or destination) is, however, an ambiguous concept in the Internet. Content distribution networks allow the same content to be accessed from multiple physical locations; a single IP address can hide a cluster of computers; or a single computer may host multiple IP addresses. Not to mention that given the potential 2^{32} addresses definable in IPv4 alone, the matrix would be rather too large. Instead, origins (and destinations) have been typically defined in terms of a prefix that specifies a group of logically related IP addresses (a subnet). The appropriate set of prefixes themselves are not trivial to define, but there is data that can be used to create a reasonable set of candidates. There are some 200,000 prefixes used currently in the Internet, so even at this level, the traffic matrix would be enormous (though likely very sparse). We could aggregate this traffic matrix in various ways (for instance, by origin and destination AS), but the most common aggregate that is used today is an Ingress-Egress traffic matrix.

An Ingress-Egress (IE) traffic matrix gives the traffic volumes flowing between particular ingress and egress points (links, routers or PoPs) in the AS under consideration. It is an aggregate of the OD traffic matrix. Typical large networks might have 10's of PoPs, or 100's of routers, and so the traffic matrix is of a more workable size. It is also the only full matrix that is observable by a single AS. An AS cannot observe traffic flows which do not cross its network, and so cannot infer traffic volumes that do not appear in its IE traffic matrix. The distinction between the IE and OD matrices is very important. As we will see below, they can have quite different characteristics.

In the following subsections we consider both models and inference methods for IE traffic matrices. The inference problems most commonly considered are highly underconstrained, and so we need to introduce side-information to solve them, and this information comes in the form of what is considered to be a likely model for the traffic, and so we will briefly describe the types of models that have been used for traffic matrices. Inferred traffic matrices are used by some of the large backbone service providers to perform network design, traffic engineering [47, 86, 87, 94], or reliability analysis [126], and commercial tools now exist for performing these tasks, e.g., [77, 19].

2.2.1. *Modeling Ingress-Egress traffic matrices.* Models for traffic matrices started with modeling the individual traffic matrix flows. The first work on traffic matrix inference suggested use of a Poisson process model [113, 105] for the arrival of packets, however, the Poisson model for Internet traffic is widely known to be false, and so later papers adopted Gaussian models where traffic volumes in each time interval were formed from a

Gaussian process [17, 124]. In such a process, typically one assumes some non-stationary mean exists, about which there are Gaussian variations, and it is interesting to consider the relative variations around the mean. A number of papers have commented on this relationship [68, 85, 88, 50], which is clearly not Poisson at the time-scales of interest.

There are also a number of papers that have noted that the non-stationary mean of Internet traffic may follow patterns [88, 60, 58, 59, 112], in particular, it is common to observe diurnal and weekly cycles in the traffic, as well as long-term trends. These seasonal, and long-term patterns exhibit variations often an order of magnitude greater than statistical fluctuations in the traffic, and so are at least as important to characterize as the variations around the mean.

Temporal models for individual flows are clearly limited — by focusing only on the temporal characteristics of flows, they fail to capture the interesting correlation structure across the traffic matrix elements. We refer to models that focus on the correlation structure of the matrices as *spatial models*. Such models first appeared in [68] in the form of the *choice models*. However, a simpler and more intuitive model for traffic matrices is the *gravity model*. The gravity model was first used in the context of Internet traffic matrices in [127], though the idea is considerably older. Gravity models take their name from Newton's law of gravitation, and are commonly used by social scientists to model the movement of people, goods or information between geographic areas [110, 81, 80]. In Newton's law of gravitation the force is proportional to the product of the masses of the two objects divided by the distance squared. Similarly, in gravity models for interactions between cities, the relative strength of the interaction might be modeled as proportional to the product of the cities' populations. A general formulation of a gravity model is given by $X_{ij} = \frac{R_i \cdot A_j}{f_{ij}}$, where X_{ij} is the matrix element representing the force from i to j; R_i represents the *repulsive* factors that are associated with leaving from i; A_j represents the *attractive* factors that are associated with going to j; and f_{ij} is a friction factor from i to j.

In network applications, gravity models have been used to model the volume of telephone calls in a network [54]. In the context of Internet TMs, we can naturally interpret X_{ij} as an OD or IE traffic matrix, the repulsion factor R_i as the volume of incoming traffic at location i, and the attractivity factor A_j as the outgoing traffic volume at location j. The friction matrix (f_{ij}) encodes the locality information specific to different source-destination pairs, however, as locality is not as large a factor in Internet traffic as in the transport of physical goods, we shall assume a common constant for the friction factors. The resulting gravity model simply states that the traffic exchanged between locations is proportional to the volumes entering and exiting at those locations.

Formally, denote the network nodes by n_i, $i = 1, \ldots, N$, and the IE traffic matrix by T, where $T(i,j)$ denotes the volume that enters the network at node n_i and exits at node n_j. Let $T^{\rm in}(i)$ and $T^{\rm out}(j)$ denote the total traffic that enters the network via node n_i, and exits the network via node n_j, respectively. The gravity model can then be computed by either of

$$T(i,j) = T^{\rm tot} \frac{T^{\rm in}(i)}{\sum_k T^{\rm in}(k)} \frac{T^{\rm out}(j)}{\sum_k T^{\rm out}(k)} = T^{\rm tot} p^{\rm in}(i) p^{\rm out}(j), \tag{2}$$

where $T^{\rm tot}$ is the total traffic across the network, and $p^{\rm in}(i)$ and $p^{\rm out}(j)$ denote the probabilities of traffic entering and exiting the network at nodes i and j respectively. Under the conservation assumption that the network is neither a source nor sink of traffic in itself $T^{\rm tot} = \sum_k T^{\rm in}(k) = \sum_k T^{\rm out}(k)$ and we can also write

$$p(i,j) = p^{\rm in}(i) p^{\rm out}(j), \tag{3}$$

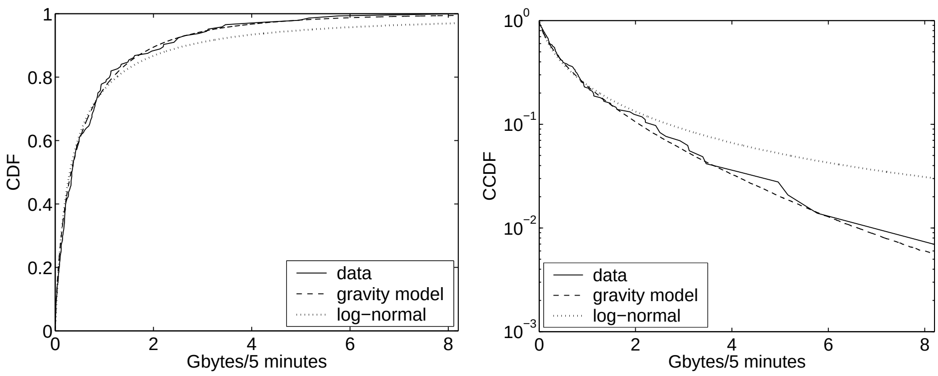
where $p(i, j)$ is the probability that a packet (or byte) enters the network at node n_i and departs at node n_j. Hence the gravity model corresponds to an assumption of independence between source and destination of the traffic. More importantly, using the above, the gravity model can be written as a matrix formed from the product of two vectors, e.g.

$$P = \mathbf{p}_{\text{in}}\mathbf{p}_{\text{out}}^T, \tag{4}$$

so by characterizing these two vectors, we obtain a reasonable characterization of the matrix.

In the form just described, the gravity model has distinct limitations. For instance, real traffic matrices may have non-constant f_{ij}, (perhaps as a result of different time-zones), and there is a problem (discussed later) with the use of such a model for IE matrices. However, the model has been shown to be quite useful because it can replicate some statistics of actual traffic matrices very well. Figure 3 shows that the Cumulative Distribution Function (CDF) and Complimentary Cumulative Distribution Function (CCDF) of real traffic matrix elements in comparison to a gravity model synthesis technique (and a simple log-normal model suggested elsewhere in the literature [76]). We can see that the distribution functions for the real traffic match a gravity model very well (see [84] for details). Of course, as noted above, simply replicating statistics is not a sufficient validation of the model. Hence, we cannot convey explanatory power from the model, but the model can nevertheless be useful in a number of applications (e.g. inference, and synthesis). Moreover, the independence assumption can be rewritten $p(i|j) = p^{\text{out}}(i)$, i.e., the conditional probability of choosing a particular exit point, given an entry point, is just the probability of choosing that exit point. In other words, traffic is homogeneous in the sense that its origin (and similarly destination) don't influence the destination (origin). Homogeneity is a reasonable initial assumption for Internet traffic.

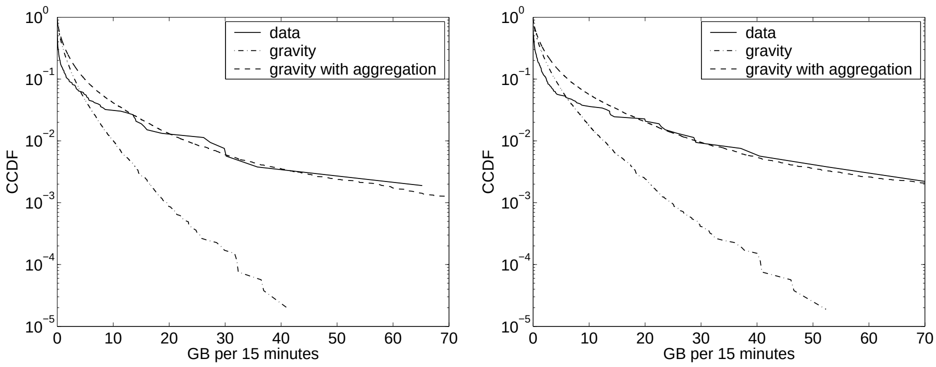
Note also some interesting features of these distributions: the traffic matrix clearly comes from a skewed distribution, the distribution follows a rough 80-20 law (80% of traffic is generated by the largest 20% of flows). Similar distributions have often been observed, and so traffic matrix work often concentrates on these larger flows, because of their relative importance. However note that the distribution is not heavy-tailed in the conventional sense that the tail follows a power-law, and in fact, the distribution has a lighter tail than the log-normal distribution.



(a) CDF (b) CCDF

FIGURE 3. A comparison between the log-normal, and the gravity model fits to the empirical Abilene data (a single 5 minute PoP-PoP TM from 00:15 on the 1st of March, 2004).

Figure 3 shows that the gravity model can lead to methods for synthesizing realistic traffic matrices for the Abilene network. In this paper we also provide new results for the GÉANT network, which provides Internet access to universities and research organizations in Europe. Results for the GÉANT network (derived from data provided in [112]) are shown in Figure 4. The synthesized gravity model (generated as for Abilene) no longer fits the traffic matrix. To understand this, we must delve a little into the GÉANT network. It is important in this network to note that nodes in GÉANT's network provide access to regional aggregation networks. However, this happens in different ways. In most cases, the regional aggregation network is independent of GÉANT, except for transit to other networks. The net impact is that GÉANT doesn't see any traffic between sub-nodes of the regional network. Typically, for a network where regional traffic does not have to come up to the backbone, the diagonal elements of the matrix will be zero and they won't match the gravity model. However, in at least one case on GÉANT, the regional network uses the GÉANT router to transit traffic within the regional network, and so this traffic does appear on GÉANT, and inflates the values of matrix elements at this node. In particular, one node in GÉANT aggregates 7 other nodes traffic. Figure 4 shows the result of generating a simple gravity model with $N + 7$ nodes but aggregating eight of these nodes together. We can see that at least for the large traffic matrix elements the fit is again good.



(a) 05:15 (b) 22:15

FIGURE 4. GÉANT traffic matrices (from August 30th 2005). In this case the simple gravity model fails to provide a usable model for the large traffic matrix elements because GÉANT has several nodes which aggregate other nodes traffic. Including one such aggregation when creating the matrices produces a better fit.

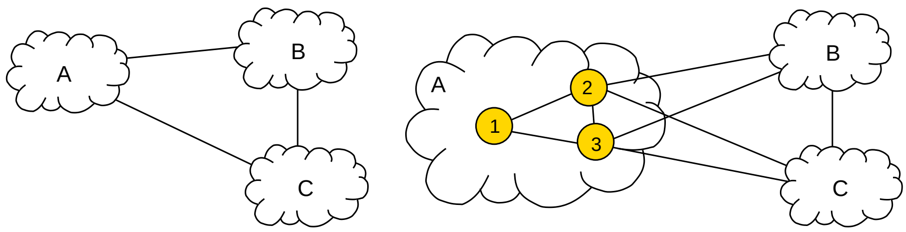
The interesting point to notice in the study of the GÉANT example is that the causality between topology and traffic doesn't just flow from traffic to topology (via network design). The network topology also has an impact on the *observed* traffic. The GÉANT example shows that to correctly build a traffic matrix, one must understand the network topology. Another way in which this is true results from the study of IE traffic matrices. As noted, these are more commonly studied than OD matrices. However, IE matrices can be distorted by inter-domain routing. We provide here a simple illustrative example of how this can commonly occur.

The Internet is made up of many connected networks. Often they interconnect at multiple points. The choice of which route to use for egress from a network can profoundly change the nature of IE traffic matrices. Typically, networks use hot-potato routing, i.e., they choose the egress point closest to the ingress point, and this results in a systematic distortion of IE traffic matrices away from the simple gravity model.

In order to understand this, let us consider a toy example consisting of three ASes, representing the three separate, but connected networks shown in Figure 5 (a). We shall assume that each connects with the other, and passes traffic between each, with the total volume passed between each following a gravity model, i.e. $T = \mathbf{t}\mathbf{t}^T/T^{\text{tot}}$, where for the sake of example, let us take $\mathbf{t} = (3,3,3)^T$, so that

$$T = \begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \tag{5}$$

The row label denotes the source, and the column the destination.



(a) A simple inter-AS network.

(b) The simple network showing the internal router level structure of A.

FIGURE 5. Example networks.

Now, let us refine the model by delving into the details of network A, which we shall assume consists of three routers as shown in Figure 5 (b). Assume that a gravity model holds between routers, and their aggregates (in this case networks B and C are still aggregates of routers), and the traffic is evenly spread between the routers of A, then

$$T' = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & B & C \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ B \\ C \end{array} & \begin{pmatrix} 1/9 & 1/9 & 1/9 & 1/3 & 1/3 \\ 1/9 & 1/9 & 1/9 & 1/3 & 1/3 \\ 1/9 & 1/9 & 1/9 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 & 1 & 1 \\ 1/3 & 1/3 & 1/3 & 1 & 1 \end{pmatrix} \end{array} \tag{6}$$

Notice that

- the matrix as a whole still follows a gravity model with $t = (1,1,1,3,3)^T$,
- the sums of appropriate submatrices still match (5).

The traffic matrices T and T' are OD traffic matrices — they specify volumes of traffic between origins and destination (whether these be routers or ASes).

In order to obtain IE traffic matrices, we need to determine the ingress and egress points of traffic on A's network. Assume that A, B and C are peers, and that shortest-AS path is used for inter-domain routing (this is not necessarily the case but we examine the simplest case for the purpose of exposition). Assume also that hot-potato routing is used internally by A and that the Interior Gateway Protocol (IGP) weights are all equal. The following exit points will be chosen for traffic originating at a router in A's network (destined for network B or C): traffic originating at router 2 will exit the network at router 2, traffic originating at router 3 will exit the network at router 3, and traffic originating at router 1 will exit the network at routers 2 and 3 for destinations B and C respectively.

We can decompose the IE traffic matrix into four components

1. *Internal traffic:* this is traffic from one router to another within A, and is given by the top-left 3×3 submatrix of T', i.e.,

$$T^{\text{internal}} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \left(\begin{array}{ccc} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{array} \right) \end{array} \tag{7}$$

2. *Traffic departing A:* is traffic that originates at a router in A (or perhaps from a customer network of A), and it is routed using hot-potato routing. In the example presented the resulting traffic looks like:

$$T^{\text{departing}} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \left(\begin{array}{ccc} 0 & 1/3 & 1/3 \\ 0 & 2/3 & 0 \\ 0 & 0 & 2/3 \end{array} \right) \end{array} \tag{8}$$

3. *Traffic coming into A:* is traffic originating in networks B and C with a destination in A. The entry points of this traffic are controlled by B and C respectively, and so, from A's point of view, this traffic is randomly distributed across the ingress links. For example, assuming an even spread, the traffic matrix would appear like

$$T^{\text{arriving}} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \left(\begin{array}{ccc} 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{array} \right) \end{array} \tag{9}$$

4. *External traffic* is traffic going between B and C (and any other networks). None of this traffic appear on A's network (A does not provide transit for its peers).

Adding up all of the traffic with the same entry and exit points we get

$$T^{IE} = \left(\begin{array}{ccc} 1/9 & 4/9 & 4/9 \\ 4/9 & 10/9 & 4/9 \\ 4/9 & 4/9 & 10/9 \end{array} \right) \tag{10}$$

which doesn't match a gravity model at all — for instance, we can see the sudden appearance of larger diagonal terms in this matrix. Thus the IE traffic matrix may not appear to be generated by a gravity model, even though the OD traffic matrix is directly generated this way. It is important to understand the difference between IE and OD traffic matrices, because each has limitations: OD traffic matrices are needed to understand or predict results of inter-domain routing changes, but are much harder to measure, whereas IE traffic matrices may be changed by modifications to routing, without any fundamental change in the users usage patterns. In essence, OD matrices are invariant under a larger group of topology (and routing) changes than IE matrices, though the latter are still useful in many applications.

Obviously the example above is a toy to illustrate the various issues involved. However, in [127] a *generalized gravity model* was proposed that incorporated these features. The paper shows that the generalized gravity model can provide a much better fit to real traffic matrices than the simpler model and that it is, in fact, equivalent to the independence assumption of the simple gravity model, conditional on the class of the ingress and egress points.

The gravity model can be generalized in other ways. For instance, it is closely related to the concept of a *random dot-product graph* [46, 91]. A random dot product graph is generated in the following manner. Start with N nodes, and generate (in some fashion) a vector at each node i, which we will denote $\mathbf{x}_i$. Then we create a dot-product matrix, i.e. we create

$$F_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j = \sum_k x_i^{(k)} x_j^{(k)}, \tag{11}$$

where $x_i^{(k)}$ is the kth element of the vector $\mathbf{x}_i$. In a dot-product graph, the probability that an edge i–j appears in the graph is given by $f(F_{ij})$, where f is some function (conventionally used to ensure that the values are probabilities, i.e. lie in the interval $[0, 1]$). The formalism is appealing because it includes large classes of random graphs (threshold graphs, Erdos-Renyi graphs, etc.), and has been suggested in social networks to model interactions between entities. For instance, A might associate with B if they have many features in common, where the features in question are recorded in the vectors $\mathbf{x}_A$ and $\mathbf{x}_B$, respectively.

However, for the case of vectors of length one (scalars) the matrix F is simply a gravity matrix. So we can see that the formalism above provides a mechanism for modeling a more general type of traffic matrix, formed by the linear combination of a series of K gravity models. Such a superposition of matrices does not have compelling motivation, but with one additional twist we can provide such. We generalize the formalism above to create *random matrix-product graphs* by taking

$$F_{ij} = \mathbf{x}_i^T Q \mathbf{x}_j, \tag{12}$$

where Q is an arbitrary (fixed) matrix. The representation would be equivalent to a dot product where Q is positive definite [46], but we do not restrict ourselves to these cases. An example of where this might be of particular use would be in the modeling of traffic generated by server-client interactions. In such interactions, the volume of traffic is often asymmetric. In particular, we might model the volume of traffic between servers and clients as proportional to the number of clients, times the number of servers. In this case, the vector $\mathbf{x}_i = (\#\text{of clients}, \#\text{of servers})$ and the matrix product in question would be

$$F_{ij} = \mathbf{x}_i^T \begin{pmatrix} 0 & \alpha \\ 1 & 0 \end{pmatrix} \mathbf{x}_j, \tag{13}$$

where α is a constant expressing the degree of asymmetry in the traffic. There is a great deal of scope for investigation of such models, and their applicability to Internet traffic matrices.

The gravity model is a purely spatial model, while the models described earlier focus on temporal characteristics of traffic. It is reasonable to consider whether spatio-temporal models can be constructed. Two trains of research exist on this topic. Firstly, [60, 58, 59] have used *principal component analysis (PCA)* to construct a spatio-temporal model for traffic. The authors break traffic into principle eigenvectors (along the time axis), the first few of which correspond to the temporal cycles in traffic (diurnal and weekly cycles), and most other correspond to Gaussian variations around these. The traffic is correlated across matrix elements by synchronizations in the periodic components (due to similarities in time-zones across the networks studied). The second approach [76] has been (effectively) to generate a probability matrix such as is discussed above, and then apply the simple temporal models above to create the actual traffic, such that the probability matrix remains constant, which has been shown to be reasonable for larger matrix elements [50].

2.2.2. *Estimating Ingress-Egress traffic matrices.* The obvious approach to obtaining an Internet traffic matrix is to measure it directly. There are various technologies such as flow-level aggregation that have been used for this purpose [42, 43]. However, in practice, this is often difficult. Routers may not support an adequate mechanism for such measurements (or suffer a performance hit when the measurements are used), and installation of stand-alone measurement devices can be costly.

On the other hand, the Simple Network Management Protocol (SNMP) is almost ubiquitously available, and has little overhead. Unfortunately, it provides only link-load measurements, not traffic matrices. Vardi's key insight was that one might be able to infer a traffic matrix from such link level measurements, and he proposed a method for doing so [113]. Vardi's general approach has similarities with tomographic techniques used in fields such as medical or seismological imaging, and hence the term Network Tomography[1]. Vardi realized the problem could be written as a linear inverse problem as follows: we observe link-load data $\mathbf{y}$ which are related by

$$\mathbf{y} = R\mathbf{x}, \tag{14}$$

to traffic matrix elements $\mathbf{x}$ written as a vector, and the routing matrix R. For a typical network the number of link measurements is $O(N)$ (for a network of N nodes), whereas the number of traffic matrix elements is $O(N^2)$ leading to a massively underconstrained linear inverse problem. There is extensive experience in solving such problems from fields as diverse as seismology, astronomy, and medical imaging, all leading to the conclusion that some sort of side information must be brought in, with the resulting accuracy being strongly influenced by the quality of the prior information. A common framework for solving such problems is regularization, where we solve the minimization problem

$$\min_{\mathbf{x}} \|\mathbf{y} - R\mathbf{x}\|_2^2 + \lambda^2 J(\mathbf{x}), \tag{15}$$

where $\|\cdot\|_2$ denotes the l^2 norm, $\lambda > 0$ is a regularization parameter, and $J(\mathbf{x})$ is a penalization functional. Approaches of this type, generally called *strategies for regularization of ill-posed problems* are more generally described in [51].

As we have seen, Vardi's method was based on side-information that traffic was Poisson (Tebaldi and West make a similar assumption [105]). Similar later approaches incorporated more realistic temporal modeling assumptions [17, 124, 115, 116]. Apart from some scalability problems in the algorithms, temporal models such as this don't take adequate account of correlations between traffic matrix elements, as shown in [68]. More recently, most algorithms for traffic matrix estimation have incorporated some type of spatial model, such as the gravity models described above [127, 128, 50, 126]. The generalized gravity model was verified to be one of the best existing technique on other networks [50, 109] (Gunnar *et al.* [50] actually showed a method using worst-case bounded priors was slightly better, but did not compare the results using a generalized gravity model which was found [128] to improve performance by several percent at least, making such a method the preferred approach). Commercial tools now exist for performing such estimates, e.g., [10, 77, 117, 19].

Of great importance is predicting how well an inference method might work for a particular network (other than those tested above). The main influence of topology in inference appears to be in determining how ill-posed the linear inverse problem is. Performance of inference appears to be approximately linear in the ratio of unknowns to measurements [128]. The higher this ratio, the worse the performance. Simple illustrative examples of this effect are a clique (a completely connected network), where (with direct routing) the traffic matrix is completely determined by the link measurements, as opposed to a star network, where the link measurements provide very little information about the traffic matrix. It is also noteworthy that the gravity model assumption is likely to work best in networks which

[1]The term has since been used for a range of related problems, such as inference of link performance from end-to-end performance measurements, and topology inference.

observe large aggregates of traffic, i.e., backbone networks. On small, local area networks, it is unlikely to be as effective.

The methods above neatly partition into temporal and spatial approaches, depending on the type of model that is used to provide side-information in the inference process. If scalability issues can be conquered, it seems that spatio-temporal models could be constructed that might include the best of both worlds.

There are a number of additional procedures one can perform to improve the traffic matrix estimates above, if more precise estimates are required. The most obvious approach is to include other measurements. In many cases, it may be difficult to measure flow-level aggregates at all of the points needed to collect a traffic matrix, but not hard to collect it at a few points. Using flow level collection at a single ingress node provides one row of a IE traffic matrix. Given a regularization approach such as described above, it is relatively easy to include additional measurements such as these, and this has been shown to produce large improvements in estimates [128, 126, 50]. Likewise, Varghese and Estan [114] have suggested other types of data that could be easily collected at a router, for instance a *local traffic matrix* giving the traffic between interfaces on the router. It was also shown in [128] that this would provide a valuable improvement in traffic matrix estimates.

An alternative to building additional measurement infrastructure is to change the existing networks in useful ways. Nucci *et al.* [75] suggest changing IGP weights in a network in order that the resulting rerouting of traffic provides measurements that would otherwise be unobserved. Using carefully chosen schedules of changes, one can create a system of equations that is no longer under-determined, and hence find a more precise estimate of the traffic matrix. Of course, such an approach assumes that the traffic matrix remains relatively constant over the period of measurement, which is not necessarily a good approximation.

Finally, it is worth noting that the above discussion has focused on point-to-point matrices. For example, we measure or infer the traffic from one ingress point to an egress point. It can be useful to consider a point-to-multipoint matrix, when considering IE matrices [43]. To understand why point-to-multipoint matrices are important, it is worth going back to the reason for studying traffic matrices in the first place. The ideal traffic matrix would be an invariant under changes to the network topology and routing. Hence, one can use the traffic matrix in determining which of a set of possible network designs (or routings) would be optimal. However, we have seen that the IE traffic matrix is not invariant to topology or routing. The point-to-multipoint traffic matrix records the amount of traffic from an ingress point to a set of egress points. The sets are chosen such that the matrix is invariant under (typical) changes in egress point. Thus the point-to-multipoint traffic matrix is more useful for planning than a simple IE traffic matrix. Zhang *et al.* [126] show how to infer point-to-multipoint traffic matrices in a very similar manner to point-to-point IE matrices, and demonstrate the utility of these new matrices.

3. **The Autonomous System-Level Internet.** A major reason for modeling the Internet's AS-level topology is to describe, characterize, and understand the logical construct that is often referred to as the "Internet Ecosystem" [74] and consists of nodes that represent *Autonomous Systems (ASes)* and (annotated) connections that indicate that the two ASes in question are in a specific type of peering relationship. Here, an autonomous system or domain denotes a group of networks operated by a single administrative entity. Inter-AS connectivity is defined in terms of pairwise logical *peering relationships.* A peering relationship between two ASes refers to a contractual business agreement between two corresponding parties (e.g., ISPs) to exchange traffic directly between them. Two ASes with an established peering relationship are physically connected by at least one direct router-level

link. The majority of peering relationships are either of the “customer-provider” type or the “peer-to-peer” type. In the former, one AS plays the role of a customer, and the other AS provides the customer with transit Internet access for a fee. In the latter, two ASes derive mutual benefits from interconnecting with each other (e.g., obtaining direct routes to the other party’s networks) and share the cost of maintaining the relationship. [2]

This AS-level ecosystem is an environment where establishing “ground truth” is notoriously difficult. The main reason is that AS-specific aspects such as physical infrastructure, traffic flows, economic aspects, or business-related data cannot in general be measured directly or are by and large proprietary in nature. Without access to direct measurements of AS-level features, the research community has been faced with the problem of identifying and collecting appropriate “surrogate” or “substitute” measurements that are publicly available or obtainable and that can be used to shed some light on the nature of this otherwise elusive AS environment. In this section, we follow closely the presentation in [25] and are concerned with two particular AS-specific aspects, *(annotated) AS connectivity maps* and *inter-AS traffic matrix* giving the traffic demand between any pair of ASes.

3.1. **(Annotated) AS-level connectivity maps.**

3.1.1. *Measurements.* Connectivity-related Internet measurements are notorious for their ambiguities, inaccuracies, and incompleteness, and many of them can at best be described as being of “limited quality.” This is true at the physical layer (see for example the discussion in Section 2) as well as at the higher layers of the protocol stack, where Internet connectivity becomes more virtual. For example, as far as measurements for inferring AS-level connectivity are concerned, network operators are generally reluctant to disclose information regarding their peering relationships with other ASes and routing policies for business reasons. Peering relationships are, in general, the result of business negotiation between two corresponding parties, or may be part of broader strategic partnership between companies. Such business-oriented peering relationships are typically protected by non-disclosure agreement. This makes it practically impossible to measure AS connectivity and the type of peering relationships directly, which in turn illustrates the need for alternative or “surrogate” measurements.

The measurements that the research community has almost exclusively relied on for inferring and modeling the Internet’s AS-level topology consist of BGP routing data sets collected by the University of Oregon Route-Views Project [89]. Here the *Border Gateway Protocol (BGP)* is the de facto standard inter-domain routing protocol deployed in today’s Internet [97]. The Oregon route server connects to dozens of operational routers belonging to commercial ISPs solely for the purpose of collecting their BGP routing data. As a result, the Oregon route-views data sets reflect AS-level connectivity, as reported by BGP, seen from a limited number of vantage points in the global Internet. Starting in Nov. 1997, the Oregon route-views data sets have been archived on a daily basis by the National Laboratory for Applied Network Research (NLANR) [71]. Presently, archives of the Oregon route-views data sets are available from `routeviews.org` [89]. In addition to the full BGP routing table snapshots, `routeviews.org` also provides daily archives of individual route updates obtained from the Oregon route server.

[2] Sometimes the term “peering relationship” is used exclusively to refer to “peer-to-peer” type relationships. “Customer-to-provider” type relationships are then referred to as “transit” relationship. Here we use the term “peering relationship” to mean *both* “peer-to-peer” type and “customer-to-provider” type relationships. Other peering arrangements (e.g., “sibling-to-sibling” relationship) do exist, but are rare and therefore not considered in this paper.

3.1.2. *Inferring AS connectivity and peering relationship.* Note that the ability to infer AS connectivity from BGP routing tables depends largely on the nature of the contract that specifies the details of an agreed-upon AS peering relationship. For example, if such a contract does not permit a given inter-AS route to be used by a third party, BGP does not advertise this information to the global Internet. Moreover, since BGP is a path-vector protocol, backup links connecting multi-homed ASes may not show up in BGP routing table snapshots. Other inaccuracies can arise either because of the dynamic nature of both ASes and peering relationships (e.g., see [22]) or because of limitations of the heuristics used to infer the type of peering relationship (e.g., see [48, 103, 118]). As a result, BGP-derived AS connectivity information is bound to provide only an inaccurate and incomplete picture of the actual AS-level Internet connectivity. However, this possibility has received surprisingly little scrutiny from the research community, even though from a BGP perspective, it should come as no surprise. After all, BGP is *not* a mechanism by which ASes distribute their connectivity. Instead, BGP is a protocol by which ASes distribute the reachability of their networks via a set of routing paths that have been chosen by other ASes in accordance with their policies. Naturally, from each AS, one can only see the subset of existing AS connections formed by these policy-influenced routes.

The first in-depth study that addresses and quantifies the degree of (in)completeness of AS connectivity maps inferred from the Oregon route-views data sets was presented by Chang *et al.* in [22] (see also [24]). By augmenting the Oregon route-views data with other publicly available and carefully sanitized data obtained from (1) full BGP table dumps from a dozen additional public route servers, (2) a selection of Internet *Looking Glass* sites that provide BGP summary information, and (3) the Internet Routing Registry or WHOIS database, Chang *et al.* made a number of important observations. First, they showed that a significant number of existing AS peering relationships remain completely hidden from most BGP routing tables. Second, the AS peering relationships with tier-1 ASes are in general more easily observed than those with non tier-1 ASes. Last but not least, there appear to be at least about 40% more AS peering relationships in the actual Internet than commonly-used BGP-derived AS maps reveal, but only about 4% more ASes. Using a much more heavy-weight, special-purpose measurement and data collection infrastructure, these findings were largely confirmed in a more recent study by Raz and Cohen [29].

3.1.3. *AS-level topology modeling and model validation.* Starting with the original observation in [40] that BGP-derived AS maps exhibit power law-type node degree distributions, popular degree-based random graph techniques for modeling the Internet's AS-level topology have included the *preferential attachment model* of Barabasi and Albert [11] and numerous variations of it, the *power-law random graph models* of Chung and Lu [33] and Aiello *et al.* [4], and very recent work described in [63] that advocates the use of the joint degree distribution to characterize BGP-derived AS connectivity. Much of this work is very traditional in the sense that it is almost exclusively descriptive in nature. However, it has also resulted in a rather weak theoretical foundation for Internet topology modeling, in general, and AS-level topology modeling, in particular (see also the discussion in Section 2). For one, in view of the studies of Chang *et al.* [22] and Raz and Cohen [29], the BGP-derived AS maps cannot be taken at face value. When taken at face value, properties of inferred AS maps that are the results of even the most sophisticated analysis of the data at hand are in general questionable, if not useless, unless they are accompanied by strong robustness results that state whether or not the observed properties are insensitive to the known ambiguities inherent in the underlying measurements. Lacking any such robustness properties, neither a modeling effort that selects a particular model based on its ability to fit the inferred AS map well, nor a model validation approach that argues for the validity of a proposed model on the basis that it is capable of reproducing certain empirically observed properties of the inferred AS map have much scientific value. Both rely on the unreasonable assumption that inaccurate measurements do not tarnish subsequent data analysis or modeling efforts, and as far as the model validation argument is concerned, it begs the questions which of the observed properties a proposed model has to match before it is deemed "valid," and how many of them. There has been an increasing awareness of the fact that two models may be identical with respect to certain properties or graph metrics, yet structurally, they can be drastically different (see for example [37, 8]).

This largely unsatisfactory situation with respect to using random graph models for describing and trying to understand the Internet's AS-level ecosystem brings up the more fundamental issue of randomness vs. design or engineering as the main forces underlying the evolution of Internet connectivity at the AS-level. Surely, deciding on whether or not to establish what type of peering relationship and with whom is largely based on economic arguments and not the outcome of chance experiments conducted by the different ASes. This then suggests a concrete alternative approach to Internet topology modeling, namely one that involves optimization of tradeoffs between multiple functional objectives of networks subject to constraints on their components, usually with an explicit source of uncertainty against which solutions must be tolerant, or robust. In this approach that has been termed HOT for *Highly Organized Tolerance* [36, 20] or *Heuristically Optimized Tradeoffs* [39], constrained optimization and robustness are the overarching themes, but models of functionality, uncertainty, component constraints, and environment are necessarily domain specific. The feasibility and success of this HOT approach in the concrete context of modeling the Internet's router-level topology is discussed in Section 2.

In the context of the Internet's AS-level topology, a HOT-based approach has been advocated and pursued by Chang *et al.* in [21, 23, 26, 24]. In particular, Chang *et al.* present a new framework for modeling the evolution of the AS-level Internet by identifying a set of concrete criteria that ASes consider either in establishing a new peering relationship or in reassessing an existing relationship. The proposed framework is intended to capture key elements in the decision processes underlying the formation of these relationships and is flexible enough to accommodate a wide range of AS-specific objectives and constraints. It includes as key ingredients AS-specific aspects such as geography (i.e., number and locality of PoPs within an AS), business rationale or operational characteristic (i.e., the primary purpose(s) behind the design, operation, and management of an AS's physical infrastructure), and traffic demands (see below). In this sense, the HOT approach emphasizes treating ASes not as generic nodes or atomic units but as geographically dispersed networks with multiple PoPs that can support a great diversity in operational characteristics, business rationale, and AS routes [70]. It also shifts the attention from the ambiguous and inaccurate BGP-based measurements to the more fundamental but not necessarily easier-to-measure AS-intrinsic aspects like geography, physical infrastructure, business model, and traffic. Note that none of these AS-specific details factor into the random graph approach to modeling the Internet's AS-level topology described above.

By focusing on the key forces at work in generating and shaping the Internet's AS-level ecosystem, the HOT approach is largely avoiding the ambiguities inherent in the BGP-based measurements and the problems that these ambiguities cause for the inferred AS maps. At the same time, the HOT perspective puts the problem of model validation in a completely new light. In fact, Chang *et al.* [21, 26] show that HOT-based models are broadly robust to changes in the underlying parameters and *perforce* yield AS connectivity maps that by and large match BGP-inferred AS connectivity with respect to most of the commonly considered graph metrics, at least to a degree where observed deviations can be fully accounted for by the known ambiguities in the underlying measurements. Note however that although – by their very construction – HOT-derived AS connectivity maps tend to fit inferred AS maps reasonably well, irrespective of the metric or property of interest, this does not mean that model validation is for free. Instead, the HOT perspective simply changes the rule of the game and requires that validation must have a different and more purposeful meaning. In fact, borrowing from the router-level topology modeling work discussed in Section 2.1, the HOT perspective argues for a more engineering-driven approach towards model validation, where the main issues are the functionality and overall performance of the system at hand as well as its robustness to the sources of uncertainty that have been identified in the HOT formulation in the first place. Naturally, functionality and actual performance measures will be domain- and system-specific, and exploring the system's functionality may involve comparisons of aspects of the actual system to those of the proposed HOT model. For example, Muhlbauer *et al.* [70] demonstrate that for routing on a modeled AS-topology to be consistent with observed AS routes, it is necessary to consider non-atomic AS structures that are capable of supporting realistic route diversity within ASes. Ensuring that this type of functionality of a proposed AS-level topology model compares well with that observed in the real world is likely to be a more profound and important aspect of model validation than matching some commonly considered graph statistics or metrics.

3.2. **Inter-AS traffic matrices.** As illustrated above, one of the key ingredients of the HOT-based approach to modeling the Internet's AS-level connectivity is the amount of traffic that is exchanged between different ASes. In particular, by examining, among other things, inter-AS traffic volumes, two potential peering partners can determine the mutual benefits associated with instantiating a peer-to-peer relationship or evaluate its cost-effectiveness. Thus, to drive this peering decision process with reasonable traffic demands, a realistic model for an inter-AS traffic matrix (i.e., a snapshot of Internet-wide traffic dynamics measured over a coarse time sale and between individual ASes) is needed. However, in stark contrast to intra-AS traffic matrix estimation (see Section 2.2), we have only incomplete knowledge of the AS-level topology, a limited understanding of inter-AS routing, and no data at all as far as "link load measurements" in the AS-level ecosystem are concerned.

3.2.1. *Difficulties measuring inter-AS traffic demands.* Unfortunately, given the highly competitive nature of today's ISP market, network operators do not make public such sensitive data as AS-wide traffic volume statistics. This makes the Internet's inter-AS traffic matrix an even more elusive object than its AS connectivity map, and as a result, research efforts to measure, model and estimate the inter-domain traffic matrix are still in their infancy. With some exceptions [41, 111, 44], most studies that require knowledge of inter-domain traffic demand typically employ extremely simple (and untested) demand models, often assuming uniform traffic demand between every pair of ASes [9, 104]. Studies such as [41, 111] that rely on traces collected from a single vantage point (typically located in some stub network) are inherently constrained in their ability to provide a global view of inter-domain traffic. Nevertheless, an analysis of their data revealed that while any given AS may exchange traffic with most of the Internet, only a small number of ASes are responsible for a large fraction of inter-domain traffic. In contrast to [41, 111], Feldmann *et al.* [44] also use proprietary server logs from a large CDN and describe a methodology for estimating the Web traffic portion of inter-AS traffic demands.

The paucity of appropriate data sets to infer inter-AS traffic demands on an Internet-wide scale has led researchers to look for "surrogate" or "substitute" measurements that are publicly available or obtainable (i.e., via measurement experiments that can be performed by anyone connected to the Internet) and that may be useful for getting a glimpse at the actual inter-AS traffic demands. The first systematic study motivating the use of possible "surrogate" measurements, designing and running Internet-wide experiments for collecting them, and exploiting them to infer inter-AS traffic demands was performed by Chang *et al.* [23, 24]. In particular, Chang *et al.* assume that to a first approximation, the traffic volume exchanged between two ASes necessarily reflects the business model of their operators. For example, an AS in the business of hosting various web and multimedia content will exhibit a very lopsided traffic profile (i.e., disproportionately heavy outbound traffic volumes). On the other hand, if two ASes are mainly in the business of providing access to residential customers and have a comparable customer basis, traffic demand between the two networks can be expected to be more symmetric. By combining a range of publicly available data sets with measurements collected from their own extensive Internet-wide experiments, they develop a "profiling" heuristic to infer the "canonical utility" of an AS's physical network as providing mainly Web hosting, residential access, or business access services. Depending on a simple high/low classification of these inferred utility values, they identify seven natural AS business models, classify each AS into one of these models, and rank the ASes within each class by their overall utility.

To illustrate the sort of data that can serve as "surrogate" measurements for estimating or inferring, say, an AS's web service utility, note that an AS that hosts popular web content or e-commerce engines as a content distribution network can be expected to carry voluminous outbound traffic and relatively little inbound traffic. To this end, Chang *et al.* [23] performed a set of network-wide experiments and collected large sets of application-layer measurements for locating popular content on the Internet. The experiments consisted of (i) obtaining a (publicly available) list of the top 10,000 most popular English keywords submitted to search engines in the years 2003-2004; (ii) using the Google Web Services API to retrieve for each submitted query seven sets (corresponding to English and six other main languages) of the top-10 most closely matched URLs; (iii) extracting web server IP addresses from these URLs by performing a carefully designed reverse-DNS lookup procedure; and (iv) mapping the IP addresses to their corresponding ASes by relying on publicly available BGP routing tables. Finally, an estimate of the byte counts of popular web content hosted by a given AS was used to define that AS's web service/hosting utility. For a more detailed description of these and other experiments to determine an AS's utility as residential Internet access provider or business access provider, see [23, 24].

3.2.2. *On modeling of inter-AS traffic matrices and model validation.* To develop an inter-AS traffic demand model that generates an inter-AS traffic matrix supported by empirical observations, Chang *et al.* [23] use a type of "gravity model" approach that is driven by the utility-based AS rankings resulting from applying their AS profiling heuristic described above. As in Section 2, this gravity model assumes that the traffic demand from AS i to j is expressed as $\frac{R_i \times A_j}{f_{ij}}$, where R_i is a repulsive factor associated with "generating" traffic at i, A_j an attractive factor associated with "absorbing" traffic at j, and f_{ij} a friction factor that "opposes" traffic from i to j. Chang *et al.* assume that for the AS-level Internet, these three factors can be expressed as functions of the empirically inferred AS rankings. In particular, representing the sum of Web traffic between i and j (where either i or j is a web hosting network) and inter-residential traffic (i.e., traffic between two residential users in i and j), the numerator captures the traffic between ASes i and j. On the other hand, the denominator can be shown to capture aspects of the service quality of the path between AS

i and AS j. Given an AS map and utility-based AS rankings, generating the corresponding inter-AS traffic matrix is relatively straightforward (see [23] for more details).

As far as model validation is concerned, the use of "surrogate" measurements clearly complicates the task, because on top of examining the validity of a proposed model, it first requires checking that the measurements in question are indeed suitable and relevant as substitutes for the otherwise unavailable data. One possible strategy for dealing with this problem is to approach regional ISPs that might be willing to provide traffic data that is sufficiently detailed to allow one to explore key features of the proposed methodology for inferring inter-AS traffic demands; e.g., the adequacy of using the "surrogate" measurements at hand and the validity of the basic traffic demand formula given by the proposed gravity model. This is the strategy followed in [23], but clearly, it leaves many (if not most) critical issues on the topic of model validation for inter-AS traffic matrices unanswered.

Also note that the inter-AS traffic matrix model proposed in [23] is not a gravity model in the strict sense, but allows for subtle interdependencies among different inter-domain traffic flows due to nature of the friction factor or denominator in the gravity model formula. On the one hand, such interdependencies may be genuine at the AS-level, where inter-dependent traffic engineering is not uncommon. On the other hand, the highly aggregated nature of quantities such as inter-domain traffic flows suggests that the gravity model assumption; that is, interactions between individual ASes are independent, seems reasonable. Validating these types of dependencies or lack thereof with appropriate data remains an open problem. In addition, we note that the modeling approach pursued by Chang *et al.* [23, 26, 24] implicitly allows for some interdependence between AS-level traffic demands and AS-level connectivity. While such an interdependence appears to be consistent with networking reality (i.e., network layout impacts traffic flow and vice versa), a more in-depth analysis of this dependence at the AS-level in both the actual Internet and its HOT-generated counterpart would be illuminating and looms as another intriguing open problem.

4. **Overlay Networks.** A theme in the previous sections has been that it is highly unlikely that generic random graph models can capture the essential features of the router-level or even AS-level connectivity in today's Internet, mainly because the lowest layers of the Internet protocol stack involving the physical infrastructure (e.g., routers, fiber-optic cables) have hard technological or economic constraints. However, the higher layers of this protocol stack define their own unique connectivity structures, and since the corresponding network topologies become by design increasingly more virtual and unconstrained, it is conceivable that certain random graph constructions could provide accurate and useful models for such *virtual* graphs as, for example, the World Wide Web (WWW) or other types of overlay networks. In the following, we discuss some examples of such virtual graphs in more detail, focusing in particular on the available measurements, dominant modeling paradigms, and issues of model validation.

4.1. **The Web graph.** The Web graph may be viewed as a directed graph where the nodes and directed edges represent Web pages and hyperlinks, respectively. In contrast to the Internet's router-level graph, the Web graph is expressively not a representation of any aspect of the Internet's physical infrastructure and is by design essentially completely unconstrained. At the same time, while networks such as the Internet's router-level or AS-level graphs are largely static or change only very slowly (over time scales on the order of weeks or months for router-level, and hours or days for AS-level graphs), the Web graph is highly dynamic, with new nodes/edges being added and existing nodes/edges deleted or changed constantly, typically over time scales on the order of minutes or seconds.

The highly dynamic nature and the enormous size have made the Web graph an interesting object to measure, analyze, and model. The measurements that have informed much of the recent research on Web graphs are based on more or less extensive crawls that provide the data to create a static representation of the Web's interconnectivity structure. For example, Broder *et al.* [15] report on two AltaVista crawls performed in May and October of 1999 that produce Web graphs with about 200 (270) million nodes and 1,500 (2100) million links; a crawl of the entire `nd.com` domain by Albert *et al.* [5] resulted in a graph with about 325,000 nodes and 1,500,000 links. Such crawls typically last a couple of weeks and the resulting graphs provide a static snapshot of the aggregate of all URLs and hyperlinks encountered during that period. One of the few large-scale crawling studies that allows for an examination of inferred Web graphs over time is by Fetterly et al. [45], who fetched over 150 million web pages per week for a period of 10 weeks in late 2002. There are currently no known Web crawlers that can account for the fact that the structure that they examine is changing underneath them so as to produce reliable input for the generation of snapshots of Web graphs over time scales that are significantly shorter than a week.

In a large body of recent work, graph properties of these static snapshots of the Web's interconnectivity structure have been examined, with special focus on power-law (or scale-free) node degree distributions (e.g., [55, 56, 5, 15]), small world property and diameter (e.g., [5, 15]), connected components (e.g., [15, 55]), bipartite cores (e.g., [52, 55]), and self-similarity [34]. This work has motivated the development of new models of the Web graph that account for the various observed properties and imitate the dynamic or evolving nature of the Web in the sense that nodes and edges can appear or disappear over time. Three of the most popular approaches are the *preferential attachment models* of Albert and Barabasi [11] or their mathematically more rigorous counterpart, the *linearized chord diagram models* of Bollobas et al. [12] (see also [30]); the *evolving copying models* of Kumar *et al.* [55] and variations thereof (e.g., see [3]; and the *growth-deletion models* of Chung and Lu [33] and Cooper *et al.* [31]. A more complete recent survey of models of the Web graph can be found in [13] (see also [14]).

The popularity of these models notwithstanding, their emphasis on reproducing various properties of inferred Web graphs and the resulting efforts to explain them in the Web context leaves room for significant improvements. For one, the crawler-based measurements that are key to inferring the Web's connectivity structure are clearly of limited quality when it comes to exploring properties of the evolutionary aspect of the Web graph and hence offer limited opportunities for validating the dynamic aspects of the proposed graph models. Moreover, being purely descriptive models, they are in general not able to provide correct physical explanations for the observed structural or temporal properties of measured Web graphs. The claim is that, in the absence of a basic understanding of the main drivers of network structure and evolution, it is difficult to identify the causal forces affecting large-scale properties of the Web graph and even more difficult to predict future trends in its evolution.

4.2. **P2P networks.** P2P systems have become increasingly popular, with many millions of simultaneous users and covering a wide range of applications from file-sharing programs like LimeWire and eMule to Internet telephony services such as Skype. A natural interpretation of a P2P network is a graph, where the peers are the vertices and where two peers are connected by an edge if and only if there is an active network connection (e.g., via TCP) between them. Like the Web, these graphs are expressively not a representation of any aspects of the Internet's physical infrastructure, except, of course, that an edge between two peers implies that they can exchange packets along some route in the Internet. However, the fact that they are neighbors in the P2P graph says nothing about how long this route is or through how many routers the packets have to travel. While typically a few orders of magnitude smaller in scale than the Web, they exhibit a similarly highly dynamic behavior, with frequent arrivals and departures of new and existing nodes and links, respectively. At the same time, many of the popular P2P systems are by design constrained in terms of how many connections (neighbors) a peer can have and how much upload/download bandwidth it can support. In this sense, P2P graphs are more like the router-level graphs discussed in Section 2.1.

Representing a hybrid between the highly dynamic and essentially unconstrained Web graph and the largely static but seriously constrained router-level graph makes P2P networks an interesting object for study. In the following, we focus mainly on the ongoing efforts in the networking community to obtain representative measurements of existing systems that will subsequently inform the modeling and model validation effort, which in turn is expected to impact the design of future P2P protocols. Concentrating on unstructured P2P systems like Kazaa, eDonkey, or Gnutella, where peers select neighbors through a predominantly random process, a common technique to capture snapshots of the P2P topology is using sampling in conjunction with a crawler [90, 83, 2, 28] and infer P2P properties from these snapshots. However, even though the situation is akin to the Web, where the structure to be explored keeps changing underneath the crawler, little or no attention has been paid in the past to the accuracy of these snapshots and whether or not they might exhibit significant bias. An obvious potential cause of bias is the temporal dynamics of these systems, whereby new peers can arrive and existing peers can depart at any time. Locating a set of peers and measuring their properties takes time, and during that time the peer constituency is likely to change, which can often lead to bias towards short-lived peers. Another reason to be concerned about bias has to do with the connectivity structure of P2P systems. As a crawler explores a given topological structure, each link it traverses is in general much more likely to lead to a high-degree peer than a low-degree peer, seriously biasing peer selection towards high-degree peers.

Relying on a much faster and more efficient Gnutella crawler to capture complete snapshots of the Gnutella network over short periods in time, Stutzbach and Rejaie [98, 99] showed that there is indeed significant bias associated with the early P2P measurement studies. In fact, they found that the bias can be so strong that commonly-made assumptions such as power-law type node degree distributions for peers or exponentially distributed peer uptimes are no longer valid. To illustrate, to a slow crawler, peers with long uptimes appear as high-degree nodes in the measured snapshot of the P2P network because many short-lived peers report them as neighbors. However, this is generally incorrect since these short-lived peers are not all present at the same time. These and similar findings have generated renewed interest in developing unbiased sampling methods as a means for obtaining representative measurements of large-scale, highly dynamic, unstructured P2P systems. Recent work by Stutzbach *et al.* [100, 101, 102] addresses this problem and develops a crawler based on a sampling technique that produces nearly unbiased samples under a variety of circumstances commonly encountered in actual P2P systems. Because of their improved accuracy, these next-generation measurements can be expected to lead to more adequate models of P2P networks, allow for a more faithful characterization of churn (i.e., the dynamics of peer participation or membership dynamics), and be more relevant for future P2P system design.

5. **Conclusions and Outlook.** Two of the critical features of the architectural design of the Internet are its vertical decomposition into layers and its horizontal decentralization across network components. As a result, there are many different facets of “network topology” and “network traffic,” and the precise meaning depends directly on one’s choice of focus.

For example, while the focus in Section 2 is on the physical layout (i.e., Layer 2) of an AS and on the corresponding intra-AS traffic matrix, Section 3 is concerned with the Internet's AS-level connectivity structure and the corresponding inter-AS traffic matrix. The picture that has emerged during the past five years of the large-scale statistical properties of these two connectivity maps strongly suggests that while they may appear deceivingly similar when viewed from the perspective of certain graph statistics or metrics (e.g., node degree distribution), their structures are often completely different and are likely shaped by very different forces and factors.

In both examples, we have seen indications of explicit or implicit correlations between network topology and traffic. After all, network topologies are designed to carry traffic, and as shown in Section 2.2, when combined with routing, the topology can have a profound impact on the IE traffic matrix of an AS. On the other hand, we assumed in Section 3.2 that – in the presence of shortest-path routing – the inter-AS traffic matrix directly influences the Internet's AS-level connectivity structure. While a fundamental understanding of these correlations is still missing, the examples in this paper suggest that routing plays a key role in illuminating and possibly exploiting the impact of traffic on topology and vice versa. A number of recent papers have started to delve more deeply into the relationships between topology, routing and traffic. For instance, Teixeira *et al.* [107, 108] have considered the impact of routing changes due to the Internet Gateway Protocol (IGP) on the IE traffic matrix, while Zhang *et al.* [126] have explored how topology-traffic correlations (e.g., between node degree and the amount of traffic seen at that node) can be seen in inference methods.

While we treated the intra-AS and inter-AS scenario in this paper by and large separately, they clearly represent two opposite ends of a multi-scale spectrum of topology/traffic examples. To illustrate, with detailed annotated maps of the physical infrastructures of individual ASes now within reach, there are natural ways of "coarse-graining" or "coarsifying" such maps to obtain less detailed representations of "network topology" that are nevertheless physically or logically meaningful. For example, one natural coarsification of the physical connectivity could represent Internet connectivity at the IP layer (Layer 3) as seen by traceroute. Coarsifying yet further could result in a PoP-level view of Internet connectivity. Finally, coarse-graining even further by collapsing all PoPs within an AS, combined with an adequate accounting and annotation of all physical links, would result in annotated AS-level maps that include such details as network ownership, capacity, PoP-level connectivity and geography, routing policies, etc. The possibility to exploit this networking-specific, multi-scale view of Internet topology for the purposes of network data representation, analysis, and visualization looms as a promising open research problem, especially when combined with estimating or inferring the different traffic matrices associated with the different "scales" and studying their multi-scale properties.

In fact, it suggests the development of an Internet-specific *Multi-Resolution Analysis* technique; that is, a structured approach to representing, analyzing, and visualizing Internet-related measurements that respects the critical design aspects of today's Internet architecture, including its dual decomposition of functionality.

Acknowledgments. Matthew Roughan's participation in this work was supported by Australian Research Council Grant DP0665427, and funding from the Defense Science and Technology Organization of Australia. Steve Uhlig's participation in this work was funded by the F.N.R.S, Belgium. David Alderson's participation in this work was supported at the California Institute of Technology (Caltech) by Boeing, AFOSR URI 49620-01-1-0365 "Architectures for Secure and Robust Distributed Infrastructures", AFOSR Award:

FA9550-05-1-0032 "Bio Inspired Networks," and Caltech's Lee Center for Advanced Networking.

The authors would like to thank the Cooperative Association for Internet Data Analysis (CAIDA) at the University of California San Diego for the use of Figure 1. We would also like to acknowledge many useful discussion with various colleagues including Yin Zhang and Miro Kraetlz.

REFERENCES

[1] D. Achlioptas, A. Clauset, D. Kempe and C. Moore, *On the Bias of Traceroute Sampling (or: Why almost every network looks like it has a power law),* in Proc. of ACM STOC 2005 (Baltimore, May 21-24).

[2] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. Huberman, *Search in power-law networks,* in Physical Review E 64(4):46135–46143, 2001.

[3] A. Adler and M. Mitzenmacher, *Towards compressing web graphs,* in Proc. Data Compression Conference, 2001.

[4] W. Aiello, F. Chung, and L. Lu, *A random graph model for power law graphs,* in Proc. ACM Symp. on Theory of Computing, pp. 171–180, 2000.

[5] R. Albert, H. Jeong, and A.-L. Barabasi, *Diameter of the World-Wide-Web,* in Nature 401, 130, 1999.

[6] R. Albert, and A.-L. Barabási. *Statistical Mechanics of Complex Networks,* In Rev. of Modern Physics (74), 2002.

[7] D. Alderson, J. Doyle, R. Govindan, and W. Willinger. *Toward an Optimization-Driven Framework for Designing and Generating Realistic Internet Topologies,* In Computer Communications Review (Proc. ACM SIGCOMM HotNets-1), Vol. 33, pp. 41–46, 2003.

[8] D. Alderson, L. Li, W. Willinger, and J. C. Doyle, *Understanding Internet topology: Principles, models, and validation,* in IEEE/ACM Transactions on Networking, 13(6):1205–1218, 2005.

[9] A. Akella, S. Chawla, A. Kannan, and S. Seshan, *A measurement-based analysis of multihoming,* in Computer Communication Review (Proc. 2003 ACM SIGCOMM), 33(4), 2003.

[10] *Arbor networks.* `http://www.arbornetworks.com/products_sp.php#TrafficRouting`.

[11] A.-L. Barabási and R. Albert. Emergence of scaling in random networks, Science 286, 509-512 (1999)

[12] B. Bollobas, O. Riordan, J. Spencer, and G. Tusnady, *The degree sequence of a scale=free random graph process,* in Random Structures and Algorithms 18, pp. 279–290, 2001.

[13] A. Bonato, *A survey of models of the web graph,* in Proc. of Combinatorial and Algorithmic Aspects of Networking, 2004.

[14] A. Bonato and J. Janssen, *Infinite limits of copying models of the web graph,* in Internet Mathematics 1(2):115–126, 2003.

[15] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tompkins, and J. Wiener, *Graph structure in the web,* in Computer Networks 33, pp. 309–320, 2000.

[16] T. Bu and D. Towsley. *On distinguishing Between Internet Power Law Topology Generators,* In Proc. IEEE INFOCOM 2002.

[17] J. Cao, D. Davis, S. V. Wiel, and B. Yu, *Time-varying network tomography,* Journal of the American Statistical Association, pp. 1063–1075, December 2000.

[18] K.L. Calvert, M. Doar, and E. Zegura., *Modeling Internet topology,* in IEEE Communications Magazine 35, pp. 160–163 (1997).

[19] *Cariden.* `http://www.cariden.com/technologies/papers.html`.

[20] J. M. Carlson and J. C. Doyle, *Highly optimized tolerance: A Mechanism for power-laws in designed systems,* in Physical Review E, 60(2):1412–1427, 1999.

[21] H. Chang, S. Jamin, and W. Willinger, *Internet connectivity at the AS-level: An optimization-driven modeling approach,* in Proc. of the ACM SIGCOMM 2003 Workshops, Karlsruhe, Germany, pp. 33–46, 2003.

[22] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, *Towards capturing representative AS-level Internet topologies,* in Computer Networks, 44(6):737–755, 2004.

[23] H. Chang, S. Jamin, Z. M. Mao, and W. Willinger, *An empirical approach to modeling inter-AS traffic matrices,* in Proc. 2005 ACM SIGCOMM Internet Measurement Conference (IMC 2005), Berkeley, CA, pp. 139–152, 2005.

[24] H. Chang, An Economic-Based Empirical Approach to Modeling the Internet's Inter-Domain Topology and Traffic Matrix, Ph.D. Thesis, University of Michigan, Ann Arbor, MI, 2006.

[25] H. Chang and W. Willinger, *Difficulties measuring the Internet's AS-level ecosystem,* in Proc. 40th Conference on Information Sciences and Systems (CISS'06), Princeton, NJ, 2006.

[26] H. Chang, S. Jamin, and W. Willinger, *To peer or not to peer: Modeling the evolution of the Internet's AS-level topology,* in Proc. of IEEE INFOCOM 2006, Barcelona, Spain, 2006.

[27] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. *The Origin of Power Laws in Internet Topologies Revisited,* In Proc. IEEE INFOCOM 2002.

[28] clip2.com, *Gnutella: To the bandwidth barrier and beyond,* 2000.

[29] R. Cohen and D. Raz, *The Internet dark matter: On the missing links in the AS connectivity map,* in Proc. of IEEE INFOCOM 2006, Barcelona, Spain, 2006.

[30] C. Cooper and A. Frieze, *On a general model of web graphs,* in Random Structures and Algorithms 22, pp. 311–335, 2003.

[31] C. Cooper, A. Frieze, and J. Vera, *Random deletions in a scale free random graph process,* in Internet Mathematics (to appear).

[32] Cooperative Association for Internet Data Analysis (CAIDA), Skitter. http://www.caida.org/tools/measurement/skitter/.

[33] F. Chung and L. Lu, *Coupling online and offline analyses for random power law graphs,* in Internet Mathematics 1(4):409–461, 2004.

[34] S. Dill, R. Kumar, K. McCurley, S. Rajagopalan, D. Sivakumar, and A. Tompkins, *Self-similarity in the web,* in The VLDB Journal, pp. 69–78, 2001.

[35] M. B. Doar. *A Better Model for Generating Test Networks,* In IEEE GLOBECOM 1996.

[36] J. C. Doyle and J. M. Carlson, *Power laws, highly optimized tolerance and generalized source coding,* in Physical Review Letters, 84(24):5656–5659, 2000.

[37] J. C. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, *The 'robust yet fragile' nature of the Internet,* in Proc. National Academy of Sciences USA, 102(41):14497–14502, 2005.

[38] P. Erdos and A. Renyi. On random graphs I Publ. Math. (Debrecen) 9 (1959), 290-297.

[39] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou, *Heuristically optimized trade-offs: A new paradigm for power laws in the Internet,* in Proc. of ICALP, 2002.

[40] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the Internet topology,* in Computer Communication Review (Proc. 1999 ACM SIGCOMM), 29(4):251–262, 1999.

[41] W. Fang and L. Peterson, *Inter-AS traffic patterns and their implications,* in Proc. of IEEE Global Internet Symposium, 1999.

[42] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford, *Netscope: Traffic engineering for IP networks,* in IEEE Network Magazine, pp. 11–19, March/April 2000.

[43] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True, *Deriving traffic demands for operational IP networks: Methodology and experience,* in IEEE/ACM Transactions on Networking, pp. 265–279, June 2001.

[44] A. Feldmann, N. Kammenhuber, O. Maennel, B. Maggs, R. D. Prisco, and R. Sundaram, *A methodology for estimating interdomain Web traffic demand,* in Proc. 2004 ACM SIGCOMM Internet Measurement Conference (IMC 2004), Taormina, Italy, p. 322–335, 2004.

[45] D. Fetterly, M. Manasse, M. Najork, and J. Wiener, *A large-scale study of the evolution of web pages,* in Software Practice and Experience. Special Issue on Web Technologies 34(2):213–237, 2004.

[46] C. Fiduccia, E. Scheinerman, A. Trenk, and J. Zito, *Dot product representations of graphs,* in Discrete Math. 181, pp. 113–138, 1998.

[47] B. Fortz, J. Rexford, and M. Thorup, *Traffic engineering with traditional IP routing protocols,* in IEEE Communications Magazine 40(10):118–124, 2002.

[48] L. Gao, *On inferring autonomous system relationships on the Internet,* in Proc. IEEE Global Internet Symposium, 2000.

[49] R. Govindan and H. Tangmunarunkit. Heuristics for Internet Map Discovery, IEEE INFOCOM 2000.

[50] A. Gunnar, M. Johansson, and T. Telkamp, *Traffic matrix estimation on a large IP backbone – a comparison on real data,* in Proc. 2004 ACM SIGCOMM Internet Measurement Conference (IMC 2004), Taormina, Italy, 2004.

[51] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, 1997.

[52] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A, Tompkins, *The web as a graph: Measurements, models, and methods,* in Proc. Inter. Conf. on Combinatorics and Computing, LNCS 1627, Springer-Verlag, 1999.

[53] L. Kleinrock and W. Naylor, *On measured behavior of the ARPA network,* in AFIS Proceedings, 1974 National Computer Conference, Vol. 43. pp. 767–780, 1974.

[54] J. Kowalski and B. Warfield, *Modeling traffic demand between nodes in a telecommunications network,* in ATNAC'95, 1995.

[55] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A, Tompkins, and E. Upfal, *Stochastic models for the web graph,* in Proc. of 41st IEEE Symp. on Foundations of Computer Science, pp. 57–65, 2000.

[56] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A, Tompkins, and E. Upfal, *The web as a graph,* in Proc. 19th ACM SIGACT-SIGMOD-AIGART Symp. Principles of Database Systems, 2002.

[57] A. Lakhina, J.W. Byers, M. Crovella, and P. Xie. *Sampling Biases in IP topology Measurements,* in Proc. IEEE INFOCOM 2003.

[58] A. Lakhina, M. Crovella, and C. Diot, *Diagnosing network-wide traffic anomalies,* in Computer Communication Review (Proc. 2004 ACM SIGCOMM), 34(4), 2004.

[59] A. Lakhina, M. Crovella, and C. Diot, *Characterization of network-wide anomalies in traffic flows,* in Proc. 2004 ACM SIGCOMM Internet Measurement Conference (IMC 2004), Taormina, Italy, October 2004.

[60] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, *Structural analysis of network traffic flows,* in Proc. ACM SIGMETRICS/Performance, 2004.

[61] L. Li, D. Alderson, W. Willinger, and J. C. Doyle, *A first-principles approach to understanding the Internet's router-level topology,* in Computer Communication Review (Proc. 2004 ACM SIGCOMM), 34(4):3–14, 2004.

[62] L. Li, D. Alderson, J. Doyle, and W. Willinger, *Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications,* In Internet Mathematics **2**, 431 (2006).

[63] P. Mahadevan, D. Krioukov, M. Fomenko, B. Huffacker, X. Dimitriopoulos, K. Claffy, and A. Vadhat, *The Internet AS-level topology: Three data sources and one definitive metric,* in Computer Communication Review, 36(1):17–26, 2006.

[64] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, *Systematic Topology Analysis and Generation Using Degree Correlations,* In Proceedings of ACM SIGCOMM 2006 (Pisa, Italy).

[65] G. Malkin. Traceroute Using an IP Option. Network Working Group, Request For Comments (RFC) 1393. `http://www.faqs.org/rfcs/rfc1393.html`

[66] B.B. Mandelbrot. Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Springer-Verlag, New York, 1997.

[67] A. Medina, I. Matta, and J. Byers. *On the Origin of Power Laws in Internet Topologies,* Computer Communications Review, Vol. 30, pp. 18–28 (2000).

[68] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, *Traffic matrix estimation: Existing techniques and new directions,* in Computer Communication Review (Proc. 2002 ACM SIGCOMM), Pittsburg, USA, August 2002.

[69] M. Mitzenmacher. *A Brief History of Generative Models for Power Law and Lognormal Distributions,* In Internet Mathematics, Vol. 1, pp. 226–249 (2003).

[70] W. Muhlbauer, A. Feldmann, O. Meannel, M. Roughan, and S. Uhlig, *Building an AS-topology model that captures route diversity,* in Computer Communication Review (Proc. 2006 ACM SIGCOMM), 36(4), 2006.

[71] National Laboratory for Applied Network Research, NLANR RouteViews archive, `http://moat.nlanr.net/Routing/rawdata`.

[72] M.E.J. Newman. *The Structure and Function of Complex Networks,* SIAM Review 45, 167-256 (2003)

[73] M.E.J. Newman. *Power laws, Pareto distributions and Zipfs law,* In Contemporary Physics 46, 323-351 (2005).

[74] W. B. Norton, *The evolution of the U.S. Internet peering ecosystem,* White Paper, 2004.

[75] A. Nucci, R. Cruz, N. Taft, and C. Diot, *Design of IGP link weights for estimation of traffic matrices,* in Prof. of IEEE INFOCOM, Hong Kong, 2004.

[76] A. Nucci, A. Sridharan, and N. Taft, *The problem of synthetically generating IP traffic matrices: Initial recommendations,* Computer Communication Review 35(3), 2005.

[77] *Opnet.* `http://www.opnet.com/products/spguru/models.html`.

[78] R. Pastor-Satorras and A. Vespignani. Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge Univ. Press, 2004.

[79] V. Paxson, *Strategies for sound Internet measurements,* in Proc. 2004 ACM SIGCOMM Internet Measurement Conference (IMC 2004), Taormina, Italy, p. 263–271, 2004.

[80] R. B. Potts and R. M. Oliver, Flows in Transportation Networks. Academic Press, 1972.

[81] P. Pyhnen, *A tentative model for the volume of trade between countries,* Weltwirtschaftliches Archive, vol. 90, pp. 93–100, 1963.

[82] M. Ripeanu, *Peer-to-peer architecture case study: Gnutella network,* in 1st IEEE Int. Conference on Peer-to-Peer Computing (P2P2001), Linkoping, Sweden, 2001.

[83] M. Ripeanu, I. Foster, and A. Iamnitchi, *Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design,* in IEEE Internet Computing 6(1), 2002.

[84] M. Roughan, *Simplifying the synthesis of Internet traffic matrices,* in Computer Communication Review 35, pp. 93–96, 2005.

[85] M. Roughan and J. Gottlieb, *Large-scale measurement and modeling of backbone Internet traffic,* in SPIE ITCOM, Boston, MA, USA, 2002.

[86] M. Roughan, M. Thorup, and Y. Zhang, *Performance of estimated traffic matrices in traffic engineering,* in Proc. of ACM SIGMETRICS, pp. 326–327, San Diego, CA, 2003.

[87] M. Roughan, M. Thorup, and Y. Zhang, *Traffic engineering with estimated traffic matrices,* in Proc. 2003 ACM SIGCOMM Internet Measurement Conference (IMC 2003), Miami Beach, FL, pp. 248–258, 2003.

[88] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang, *Experience in measuring Internet backbone traffic variability: Models, metrics, measurements and meaning,* in Proc. of the International Teletraffic Congress (ITC-18), pp. 221–230, 2003.

[89] University of Oregon Route Views Project, `http://www.routeviews.org`.

[90] S. Saroiu, P. K. Gimmadi, and S. D. Gribble, *A measurement study of peer-to-peer file sharing systems,* in Proc. Multimedia Computing and Networking (MMCN), 2002.

[91] E. Scheinerman, *Random dot product graphs,* `http://engineering.jhu.edu/~nuggets/random-dot-products-graphs/`.

[92] Y. Shavitt and E. Shir. *DIMES: Let the Internet Measure Itself,* ACM SIGCOMM Computer Communication Review, 35(5):71–74.

[93] G. Siganos, M. Faloutsos, P. Faloutsos, C. Faloutsos. *Power-laws and the AS-level Internet topology,* in IEEE/ACM Transactions on Networking 11: 514–524 (2003).

[94] M. Söderqvist and A. Gunnar, *Performance of traffic engineering using estimated traffic matrices,* in Proc. of Radio Sciences and Communication RVK'05, Linköping, Sweden, June 2005.

[95] N. Spring, M. Dontcheva, M. Rodrig, and D. Wetherall. *How to Resolve IP Aliases,* UW CSE Tech. Report 04-05-04, 2004.

[96] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. *Measuring ISP Topologies with Rocketfuel,* IEEE/ACM Transactions on Networking 12: 2–16 (2004).

[97] J. W. Stewart, BGP4: Inter-Domain Routing in the Internet, Addison-Wesley, Boston, 1999.

[98] D. Stutzbach and R. Rejaie, *Capturing accurate snapshots of the Gnutella network,* in Proc. Global Internet Symposium, pp. 127–132, Miami, FL, 2005.

[99] D. Stutzbach and R. Rejaie, *Evaluating the accuracy of captured snapshots by peer-to-peer crawlers,* in Proc. Passive and Active Measurement Workshop (PAM 2005), LNCS 3431, pp. 356–360, Springer-Verlag, 2005.

[100] D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, and W. Willinger, *Sampling techniques for large, dynamic graphs,* in Proc. Global Internet Symposium, Barcelona, Spain, 2006.

[101] D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, and W. Willinger, *On unbiased sampling for unstructured peer-to-peer networks,* in Proc. 2006 ACM SIGCOMM Internet Measurement Conference (IMC 2006), Rio de Janeiro, Brazil, 2006.

[102] D. Stutzbach and R. Rejaie, *Understanding churn in peer-to-peer networks,* in Proc. 2006 ACM SIGCOMM Internet Measurement Conference (IMC 2006), Rio de Janeiro, Brazil, 2006.

[103] L. Subramaninan, S. Agarwal, J. Rexford, and R. H. Katz, *Characterizing the Internet hierarchy from multiple vantage points,* in Proc. IEEE INFOCOM, 2002.

[104] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, *Network topology generators: Degree-based vs. structural,* in Computer Communication Review (Proc. 2002 ACM SIGCOMM), 32(4), 2002.

[105] C. Tebaldi and M. West, *Bayesian inference on network traffic using link count data,* J. Amer. Statist. Assoc. 93(442), pp. 557–576, 1998.

[106] R. Teixeira, K. Marzullo, S. Savage, and G.M. Voelker. *In Search of Path Diversity in ISP Networks,* In Proc. IMC'03, Miami Beach, Florida. October 27-29, 2003.

[107] R. Teixeira, T. Griffin, A. Shaikh, and G. Voelker, *Network sensitivity to hot-potato disruptions,* in Computer Communication Review (Proc. 2004 ACM SIGCOMM), 2004.

[108] R. Teixeira, N. Duffield, J. Rexford, and M. Roughan, *Traffic matrix reloaded: Impact of routing changes,* in Workshop on Passive and Active Measurements (PAM), 2005.

[109] T. Telkamp and S. Schnitter, *Best practices for determining the traffic matrix in IP networks,* in Apricot, Kyoto, Japan, 2005.

[110] J. Tinbergen, *Shaping the world economy: Suggestions for an international economic policy.* The Twentieth Century Fund, 1962.

[111] S. Uhlig, O. Bonaventure, V. Magnin, C. Rapier, and L. Deri, *Implications of the topological properties of Internet traffic on traffic engineering,* in Proc. of ACM Symposium on Applied Computing, 2004.

[112] S. Uhlig, B. Quoitin, S. Balon, and J. Lepropre, *Providing public intradomain traffic matrices to the research community,* in Computer Communication Review 36, pp. 83–86, 2006.

[113] Y. Vardi, *Network tomography,* J. Amer. Statist. Assoc., March 1996.

[114] G. Varghese and C. Estan, *The measurement manifesto,* in 2nd Workshop on Hot Topics in Networks (HotNets-II), 2003.

[115] S. Vaton and A. Gravey, *Iterative Bayesian estimation of network traffic matrices in the case of bursty flows,* in Proc. 2002 ACM SIGCOMM Internet Measurement Workshop, 2002.

[116] S. Vaton and A. Gravey, *Network tomography: an iterative Bayesian analysis,* in 18th ITC, Berlin, Germany, 2003.

[117] *Wandl.* `http://www.wandl.com/html/news/content/prod_041103.cfm`.

[118] H. Wang, H. Xie, Y. R. Yang, E. Li, Y. Liu, and A. Silberschatz, *Stable egress route selection for interdomain traffic engineering: Model and analysis,* in Proc. IEEE ICNP, 2005.

[119] B.M. Waxman. *Routing of multipoint connections,* IEEE Jour. of Selected Areas in Comm., Vol. 6, No. 9, 1988.

[120] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, *Scaling phenomena in the Internet: Critically examining criticality,* in Proc. National Academy of Sciences **99**, suppl. 1, pp. 2573–2580, 2002.

[121] W. Willinger and J. Doyle, *Robustness and the Internet: Design and evolution,* in Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies, E. Jen, (Ed.), SFI Lecture Note Series, Oxford University Press, pp. 231–272, 2004.

[122] W. Willinger, D. Alderson, J. Doyle, and L. Li, *A pragmatic approach to dealing with high variability in network measurements,* In Proc. ACM SIGCOMM Internet Measurement Conference IMC 2004 (2004).

[123] W. Willinger, D. Alderson, J.C. Doyle, and L.Li, *More* Normal *Than Normal: Scaling Distributions and Complex Systems,* In Proceedings of the 2004 Winter Simulation Conference R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, eds. (2004)

[124] B. Yu, *Maximum pseudo likelihood estimation in network tomography,* in NISS Internet Tomography Technical Day, Research Triangle Park, 2003.

[125] E. Zegura, K.L. Calvert, and M.J. Donahoo. *A quantitative comparison of graph-based models for Internet topology,* in IEEE/ACM Transactions on Networking, Vol. 5 (1997).

[126] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, *Estimating point-to-point and point-to-multipoint traffic matrices: An information-theoretic approach,* in IEEE/ACM Transactions on Networking 13, pp. 947–960, 2005.

[127] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, *Fast accurate computation of large-scale IP traffic matrices from link loads,* in Proc. ACM SIGMETRICS, San Diego, CA, pp. 206–217, June 2003.

[128] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, *An information-theoretic approach to traffic matrix estimation,* in Computer Communication Review (Proc. 2003 ACM SIGCOMM), Karlsruhe, Germany, pp. 301–312, August 2003.

Received for publication September 2006.

E-mail address: `dlalders@nps.edu`
E-mail address: `hschang@eecs.umich.edu`
E-mail address: `matthew.roughan@adelaide.edu.au`
E-mail address: `suh@info.ucl.ac.be`
E-mail address: `walter@research.att.com`